(12)

Sponsored by

Defense Advanced Research Project Agency (DOD)

DARPA Order No. 3894

under Contract No. MDA-903-81-C-0320 issued by

Department of Army, Defense Supply Service - Washington

Washington, D.C. 20310

FINAL TECHNICAL REPORT FOR THE PERIOD 1 JUNE 1981 TO 31 AUGUST 1982

ENHANCING PERCEPTIBILITY OF BARELY PERCEPTIBLE TARGETS

Naomi Weisstein
Department of Psychology
State University of New York at Buffalo
4230 Ridge Lea Road
Buffalo, New York 14226

The views and conclusions contained in this document are those of the author and should not be interpreted as representing the official policy, either expressed or implied, of the Defense Advanced Research Projects Agency or the U. S. Government.

23 December 1982

SUMMARY

We have discovered a class of powerful procedures for improving perception of targets obscured in dim, briefly-flashed, or noisy images. These procedures require no precise knowledge of where the targets appear in the image nor are they dependent on what the targets look like. They involve adding particular kinds of spatial and temporal contexts to the obscured image. Perception then improves strikingly.

In the previous contract year, we were able to improve perceptual acurracy by as much as 20% to 100% by 1) flickering and 2) moving images and by adding contours that 3) divided the image into figure and ground regions or that 4) made the image appear three-dimensional. During the current contract year we added two new procedures to our list. 5) Simply by connecting their endpoints so as to make a more coherent pattern, we increased the apparent brightness of a display of lines by as much as 2 c/m2, even though the physical brightness of the display remained constant. 6) Simply by flickering neighboring regions so as to stimulate sensory channels linked to the perception of a background, a target image region could be made to jump out in apparent depth by as much as 5 cm in front of the neighboring regions. We also expanded our previous techniques, showing that 7) <u>reaction time</u> for a diagonal line segment is twice as fast when auxiliary lines are added that combine with it to yield perception of a three-dimensional object. 8) At certain rates of flicker, <u>absolute</u> accuracy is higher than for the same image when it is stationary (previously we had found that <u>relative</u> accuracy for more meaningful patterns is better at certain rates of flicker). Finally, we further explored 9) our previous "pixel flicker" techniques (in which we were able to double perceptual accuracy by flickering individual picture elements, moving, or adding meaningful contexts to regions in a digitized noisy photograph). Working with a set of slides supplied by DARPA (in which, unlike previous investigations, we had no advance knowledge of the content of the images or the type of distorting noise), the techniques rendered the hidden contents of the images visible.

A major virtue of our enhancement procedures is that targets do not have to be restricted to a fixed location nor does the exact location have to be known or discovered beforehand. This is obvious in the case of temporal manipulations, but it applies in general to our spatial manipulations as well. Contours that make an image region appear three-dimensional or that divide an image into figure and ground can be placed within a general area, or moved from region to region as the task requires. We found a similar lack of constraint on the types of images amenable to our enhancement procedures. Our procedures worked with such dissimilar images as randomly placed dots and short vectors, synthetic aperture radar images, small diagonal line segments, fragmented forms, and digitized photographs of faces, roadways, tanks and trucks obscured by various types of noise.

The facilitatory treatments described above may seem diverse and also paradoxical, since each adds noise or seemingly irrelevant detail to the image. In fact, however, they were all developed from a single basic principle of visual system functioning and their effectiveness apparently stems from this principle. The visual system tends to construe patterns as pictorially meaningful, and when it does, peripheral sensory mechanisms sharpen and clarify their response. Each of our manipulations added spatial and/or temporal contexts that imparted meaning to an image or that augmented the effectiveness of mechanisms for extracting meaning.

Our work on this contract suggests that we can develop a large, varied, and flexible stockpile of image enhancement techniques. There are many different ways of imparting visual meaning to an image. There are many different image types amenable to enhancement. Our findings may prove applicable to such disparate situations as

cont

picking up key words or sentences in documents, spotting faces in a crowd, identifying objects in aerial reconnaissance, surveillance and badly damaged photographs, recognizing marginal video and facsimile transmissions, and interpreting camouflaged images.

ENHANCING BARELY PERCEPTIBLE TARGETS
DETAILED REPORT

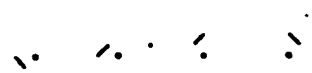
During the contract year, we added to and improved upon our previous stockpile of image-enhancing techniques. These showed that we can increase perceptual accuracy by 20% to 100% by adding certain temporal contexts--flickering and moving image regions--and/or by adding certain spatial contexts--additional lines that divide an image into figure and ground regions or that make a target within an image region appear three-dimensional. The added contexts impart meaning to an image or augment mechanisms for extracting meaning from an image. The new techniques and improvements on previous techniques that we developed this year are described, after a brief background, below.

BACKGROUND

Object Superiority

The original discovery which led to the research on this contract was the finding (Weisstein & Harris, 1974) that we could dramatically enhance observers' performance on the simple task of identifying which one of four barely perceptible, briefly flashed line segments was present

Figure 1



by adding to those line segments a fixed set of auxiliary lines:

Figure 2

which when combined with each target line, yielded perception of a unified, three-dimensional object.

Figure 3

Not just any auxiliary lines could be added, however. The auxiliary pattern had to combine with the diagonal target lines to yield perception of an object. If the auxiliary pattern combined with the target lines to yield flatter, less object-like patterns such as in the figure below, accuracy was not enhanced.

Figure 4

Since then, we have confirmed that the key factor in enhancement is imparting objectness or, more generally, meaning to the image. When the image looks meaningful, visual response becomes more accurate.

Fast Visual Response

One of our most striking further discoveries was that visual response was also <u>faster</u> to the more object-like patterns than to the flatter, less connected designs. In our initial work for DARPA, we explored whether this faster response might be utilized for image enhancement through what we termed "temporal filtering". The idea of the temporal filter was to flicker or move image elements at rates where only fast responses, i.e., those to the more meaningful parts of the image, could reach perceptual threshold. The filter worked: When we flickered or moved image elements so as to bring out pictorial meaning and to suppress slower temporal response (Genter & Weisstein, 1980a, b; Brown, Weisstein & Genter, 1981), perception improved dramatically for such diverse images as digitized photographs obscured by noise, random dots slightly brighter than their neighbors, and fragmented images partially blocked by a horizontal occluding region.

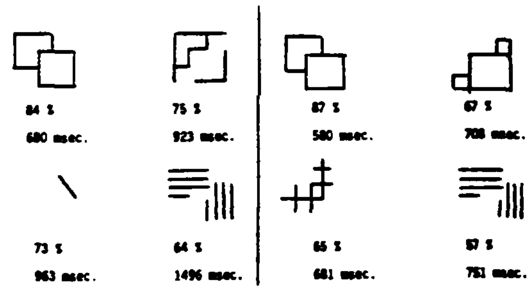
WORK DONE DURING THIS CONTRACT YEAR

Reaction Time is Faster to an Object-like Pattern

This year, we explored whether the faster visual response we had found in our temporal filtering would also be found in overall response time. It was:

Using the object-superiority design described above, but presenting only the outer two of the four target lines, we obtained observers' accuracies and reaction times to patterns varying in perceived depth and connectedness. We found that reaction time was fastest for the pattern judged highest in perceived depth and connectedness. (In one experiment, reaction time was twice as fast to the object-like pattern compared to the flatter, unconnected design.) The figure shows the results of two experiments using different displays. The first number under each figure is per cent correct; the second number is latency (Wong & Weisstein, in prep.).

Figure 5



Experiment 1: (Mean of 4 subjects) CRT vector graphics display

Experiment 2: (Mean of 7 subjects) Video display

Enhancement Implications

This finding has major implications. Previously, it was thought that there is a tradeoff between speed and accuracy: the faster one responds, the less accurate the response. Our finding that, for meaningful patterns, both reaction time and accuracy increase together is extremely relevant to military situations were responses must be both rapid and correct.

Signal Integration Functions

We further explored fast visual response using a measure designed to yield maximum detail about time course. This is the forced response method of speed-accuracy tradeoff which allows us to seperate the latency, rise-time, and amplitude of the response. The results (Wong & Weisstein, in prep.) show that responses to more object-like patterns have a higher asymptotic accuracy, faster rise-time, but longer initial latency. The faster rise-time for the more visually meaningful

Flickering a Pattern Enhances Absolute Accuracy

This year we found that certain rates of flicker not only increase the difference in accuracy for more vs. less meaningful patterns (the basis of the "temporal filter") but also can produce an increase in absolute accuracy. The flicker rate at which this absolute increase occurs is different for the more vs. the less meaningful patterns (Moravec, Ralston & Weisstein, in prep.). Temporal filtering can thus both make the more meaningful patterns more absolutely perceptible and also more distinguishable from their backgrounds.

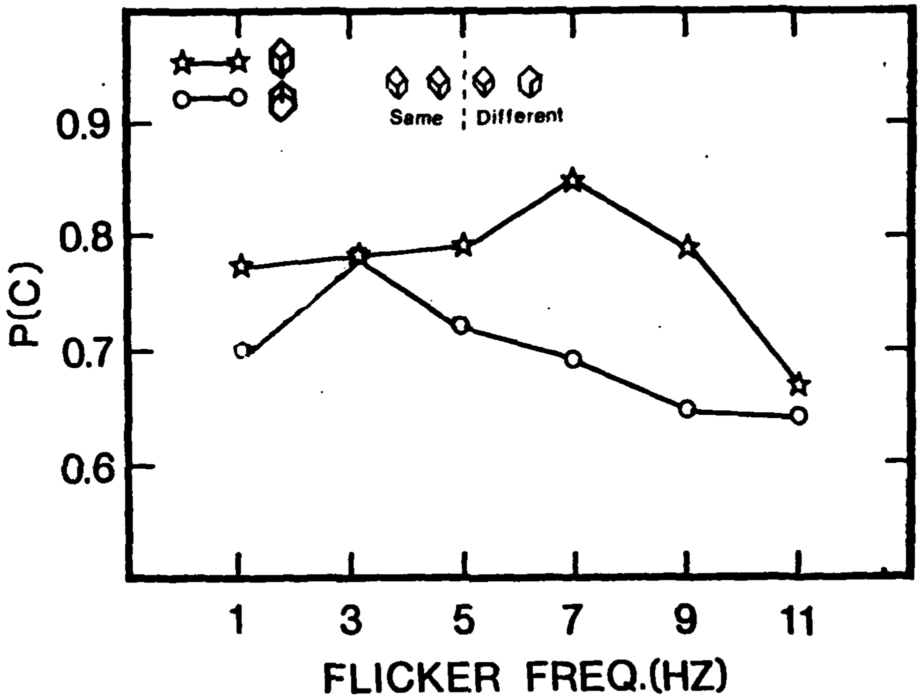


Figure 7. Mean of 20 subjects. Judgments are always between the same contexts with an internal line missing or present.

Flicker-induced Depth

In the previous year we found that with the Rubin reversible goblet-faces picture, sharp targets were detected better against whatever region of the picture was momentarily seen as figure and blurred targets were detected best against whatever region of the picture was momentarily seen as ground (Wong & Weisstein, 1982 and Note 1). If different ranges of spatial frequency are associated with figure vs. ground, then different temporal frequency ranges should also distinguish the two percepts. We flickered parts of images leaving adjacent parts stationary, and found that, as predicted, observers saw the stationary regions as figures standing as much as 5cm in front of the flickering regions. This flicker-induced depth was found with a variety of images including ERIM synthetic aperture radar images, fields of random dots, and horizontal and vertical line segments. In one experiment, we obtained spatial and temporal tuning functions using "gratings" (Figure 7) composed of alternate flickering and non-flickering random-dot bars. These functions were found to be similar to those of a "transient mechanism", with depth best at high temporal (6.3 Hz.-8.3 Hz.) and low spatial (.37 c/deg.-.74 c/deg.) frequencies (Wong, & Weisstein, 1982 and Note 2).

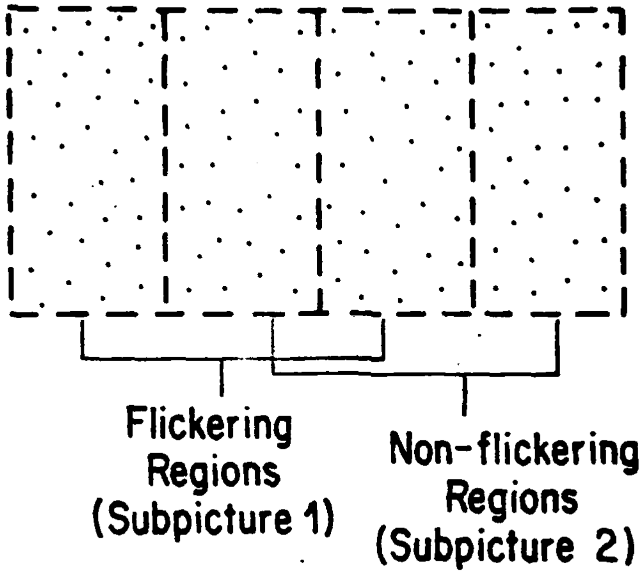


Figure 8

The "in front of" effect does not depend on the stationary regions being brighter than the flickering ones, since they are seen in front even when the average luminance of the flickering regions is twice that of the stationary ones (Figure 9, panel F:

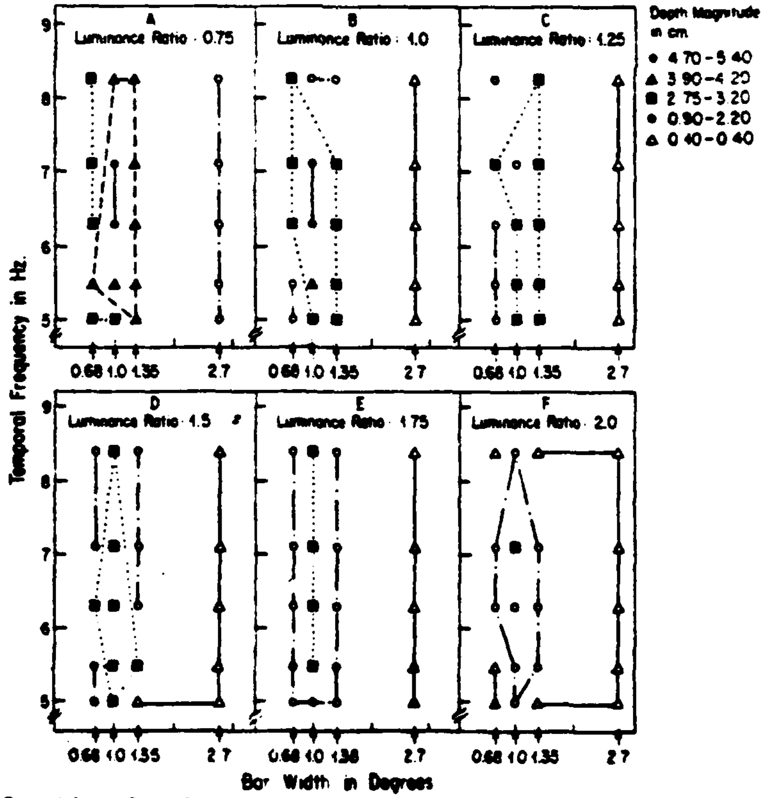


Figure 9. Tuning function for depth segregation induced by flicker. Contours connect bar width and temporal frequency loci where similar amount of depth is perceived.

Enhancement Implications

We think this technique is especially promising as a fast, flexible way to arbitrarily introduce figure and ground into desired regions of an image. During the previous contract period we had found two to three times better detection of targets appearing perceptually as a figure in front of a background. We therefore expect that this newly discovered flicker technique will allow us to greatly improve detection for targets well hidden or camouflaged within specified regions (for example, in the synthetic aperture radar images supplied to us this fall by ERIM, detection of a missile site within an orchard).

Increasing Apparent Brightness

In the previous contract year we concentrated on improving target discriminability and accuracy. This year, we also explored ease of perception, finding that we could increase the apparent brightness of a line drawing by as much as 2 cd/m2 (while keeping the physical brightness, of course, constant) merely by connecting the end points of a set of lines so that they defined a contour or set of contours rather than simply appearing as unstructured texture (Walters & Weisstein, 1982a, b). Using a balanced constant-stimuli paradigm and presenting patterns for 500 msec, 38 subjets all perceived patterns with the lines all connected to each other as being brighter than patterns with the same number of horizontal, vertical, and/or diagonal lines, but more free terminators. Differences in apparent brightness were found not only for briefly flashed displays but also when unlimited inspection time was allowed.

Enhancement Implications

Procedures to connect end points may prove useful for a variety of CRT displays where the physical brightness of the display cannot be changed easily and where a number of unrelated targets, lines, etc. must be processed.

Deciphering Images of Unknown Objects

In previous work, we were able to as much as double perceptual accuracy by flickering, moving, or adding meaninful contexts to individual picture elements in a digitized noisy photograph. Although the technique worked with a variety of images (faces, tanks, guns, convoys, etc.) in all of these cases, the content of the image and the type of obscuring noise were known in advance. During this contract period, we have been working with a set of digitized slides (supplied by DARPA System Sciences Division) without any advance knowledge of the content of the images or the type of distorting noise. A combination of our enhancement techniques has successfully rendered the hidden contents of the image visible. This greatly increases our confidence that our general research approach to image enhancement is highly flexible, and will be able to recover information from images where almost nothing is known about what is being looked for or how the image has been degraded.

SUMMARY OF WHY WE FOUND WHAT WE DID ON THIS CONTRACT

The new methods we have found for enhancing perception are illustrations of a distinctive approach to image enhancement. The approach, stated briefly, is to improve an obscured or noisy image by adding particular kinds of spatial and temporal contexts to it.

This approach runs counter to standard image enhancement procedures because it lowers, rather than raises, signal-to-noise ratios. For instance, in some of the examples above, temporal dynamics--flicker, motion, a sequence of changing images--are added to an obscured or difficult image. Adding such temporal complications to an already barely visible image might be expected to interfere with identification. But surprisingly, the correct temporal manipulation leads to the emergence of the image or its parts in sharper perceptual clarity.

Why would adding certain types of temporal and spatial 'noise' to an already obscured signal improve its clarity? The answer lies in a general principle of visual response that has emerged from our investigations. It has long been suspected (by the Gestalt psychologists and even earlier investigators--e.g., see review by Hochberg, 1979) that the visual system is particularly responsive to meaningful objects and events. Our research has found such enhanced responsiveness. When a pattern is seen as three-dimensional or otherwise pictorially meaningful, sensory mechanisms sharpen and amplify their response. Visual system response becomes *stronger and faster*. The efficiency of temporal integration increases, thresholds go down, and the number and type of active mechanisms increases. Increased activity even extends to mechanisms that do not receive direct retinal stimulation: observers see missing parts of forms in regions that are actually receiving uniform stimulation. In seeking meaning, the visual system performs its own internal image enhancement--extracting coherent images and rejecting noise.

Building on this discovery that meaning is a primary ingredient of seeing an image clearly, our approach makes use of spatial and temporal contexts that help the visual system pick up meaning in the image. Two stages are involved. First, we impose temporal and spatial pictorial meaning on an image. Second, we amplify the output of those sensory mechanisms that change their activity in response to meaning.

In the previous contract year our aim was to see the extent to which our approach would work both inside and outside specific laboratory situations, and whether it would be robust, effective, and generally applicable. To test this, we studied a range of phenomena and types of images, such as "pixel" flicker of faces, motion-in-depth of random dots, object completion of fragmented forms, and figure/ground effects on line segments. We looked for big effects--20-100% improvement in perceptual accuracy--and we found them.

This year we further explored and expanded our stockpile of enhancing techniques, showing dramatically faster overall response speed for meaningful images, absolute enhancement by flicker, perceptual improvement of images degraded by unknown noise, flicker-induced depth, and apparent brightness enhancement through connectivity.

Our efforts during the two years of this contract have thus been most successful. Our procedures are working and the effects are big. Contingent upon future funding, we plan to develop these effects into working image enhancement procedures, determining their range, limits and optimum parameters, showing how, why, and for what applications they work. Our success so far leads us to believe that we can generate a battery of fast, simple and effective procedures that have the capability to dramatically enhance images or parts of images in an extensive range of applications.

BIBLIOGRAPHY

* Indicates that the paper is included along with this report.
+ Indicates that the paper is already on file at DARPA.

Reference Notes

* 1. Wong, E., and Weisstein, N. Sharp targets are detected better against a figure and blurred targets are detected better against a ground. JEP: Human Perception and Performance, In press.

* 2. Wong, E., and Weisstein, N. Flicker induces depth: Spatial and temporal factors in the perceptual segregation of flickering and non-flickering regions in depth. Perception & Psychophysics, submitted.

References

+ Brown, J., Weisstein, N., and Genter, C. R. II. Fragmented 'street' figures can produce flickering phantom contours, moving phantom contours, and a phantom motion aftereffect. Investigative Ophthalmology and Visual Science, 20, 1981 (Abs).

+ Genter, C. R. II, and Weisstein, N. Pixel flicker: A new form of object-superiority? Optical Society of America: Recent Advances in Vision, April, 1980a (Abs).

+ Genter, C. R. II, and Weisstein, N. Perceived depth and connectedness affect flicker frequency functions. Investigative Ophthalmology and Visual Science, 19, 1980b (Abs).

Hochberg, J. Sensation and perception. In Elliot Hearst (Ed.), The first century of experimental psychology. Hillsdale, N. J.: Lawrence Erlbaum, 1979.

Walters, D., and Weisstein, N. Perceived brightness is influenced by the global structure of line drawings. Investigative Ophthalmology and Visual Science, 21, 1982a (Abs).

Walters, D., and Weisstein, N. Perceived brightness is a function of line length and perceived connectivity. Bulletin of the Psychonomic Society, 15, 1982b (Abs).

+ Weisstein, N., and Harris, C. S. Visual detection of line segments: An object-superiority effect. Science, 186, 1974, 752-755.

* Wong, E., and Weisstein, N. A new perceptual context-superiority effect: Line segments are more visible against a figure than against a ground. Science, 1982, 218, 587-589.

Wong, E., and Weisstein, N. Flicker induces depth. Journal of the Optical Society of America, 1982 (Abs).

Reprint Series
5 November 1982, Volume 218, pp. 587–589

SCIENCE

A New Perceptual Context-Superiority Effect: Line Segments Are More Visible Against a Figure than Against a Ground

Eva Wong and Naomi Weisstein

A New Perceptual Context-Superiority Effect: Line Segments Are More Visible Against a Figure than Against a Ground

Abstract. *Context, specifically the perceived figure or ground of an ambiguous form that surrounds a diagonal line segment, can influence the discrimination of that line segment even though the physical attributes of the context remain the same during figure-ground reversals. When the line segment was flashed on a region of the form seen as figure, discrimination was twice as accurate as when the line segment was flashed in isolation, and it was at least three times as accurate as when the line segment was flashed on that same region seen as ground.*

A barely visible, briefly flashed line segment is discriminated with greater accuracy when it is part of a pattern that looks like an object than when it is flashed alone or when it is part of a pattern that appears to be a random collection of lines (*1*). A letter is typically identified better when it is presented as part of a pronounceable word than when it is flashed among an unpronounceable string of letters or alone (*2*). And an object is better recognized when it is part of a coherent scene than when it is flashed in a scene whose parts have been jumbled (*3*). These object, word, and scene superiority effects can all be classified more generally as "context effects" in perception. Such context effects show that perceptual variables influence task performance quite apart from the physical aspects of the stimuli.

We now report effects of context that are entirely perceptual. Visual discrimination is dramatically enhanced when line segments are flashed in a region that is perceived as figure. Discrimination is substantially degraded when the same region is seen as ground even though the physical stimulus remains identical throughout figure-ground reversals.

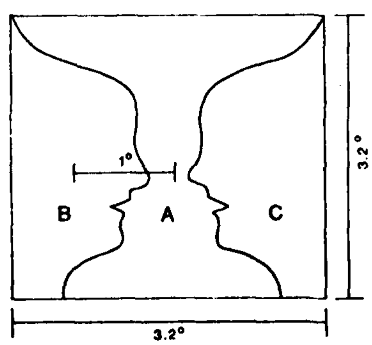


Fig. 1. Reversible face-vase figure (*4*).

In our experiment, we chose Rubin's face-vase reversible figure as the context stimulus (Fig. 1) (*4*). If one fixates at *A*, the perception of two identical faces, one on each side of the central region, alternates with the perception of a vase in the middle of the figure. When the central region is perceived as a vase (or figure) the surrounding regions become a background (ground) with no definite shape. Conversely, when the surrounding regions are seen as two faces, the central region loses its figural identity and assumes the characteristic of a formless background. The common boundary contour shared by the central and flanking regions seems to belong to the region seen as figure (*4*). In this stimulus, local and global environments, spatial frequency and phase, in fact, all the physical aspects of the stimulus are identical whether a region is seen as figure or as ground. Only the perception varies.

Our experiment compared observers' ability to identify the direction of tilt of a test line that was flashed within a given region of Fig. 1 when that region was perceived as figure or when seen as ground. The context pattern occupied a region 3.2° by 3.2° with a dim fixation point located at the center. The target was a line 0.9° long and 0.06° wide. On a

0036-8075/82/1105-0587$01.00/0

Table 1. Mean d' ($\pm$ standard deviations) across all conditions. A three-way repeated-measures analysis of variance: differences among observation conditions [$F(2, 8) = 13.86$, $P < .01$]; difference between viewing at fixation and viewing 1° left or right of fixation [$F(1, 4) = 8.64$, $P < .05$]; no interactions were significant.

Observation condition of target	Viewing at fixation: At luminance threshold, above tilt threshold	Viewing at fixation: At tilt threshold, above luminance threshold	Viewing 1° left or right of fixation: At luminance threshold, above tilt threshold	Viewing 1° left or right of fixation: At tilt threshold, above luminance threshold
On figure	1.73 ± 0.25	1.55 ± 0.21	1.10 ± 0.35	0.96 ± 0.15
Alone	0.82 ± 0.20	0.69 ± 0.15	0.42 ± 0.18	0.41 ± 0.15
On ground	0.59 ± 0.19	0.55 ± 0.12	0.26 ± 0.15	0.18 ± 0.11

given trial, this line segment, tilted left or right, was flashed for 20 msec at one of three positions (*A*, *B*, or *C* in Fig. 1). The target line was always 0.5° from the contours making up the context pattern.

A computer (PDP-11) with a graphics display processor (GT-40) generated the stimuli, controlled the experiment, and collected and analyzed the data. Observers viewed the display monocularly and were instructed to fixate the dim fixation point during each trial. Since observers tend to fixate a figure, they might direct their gaze to the flanking regions when faces are perceived. If the target appeared in the central region (or ground) while observers were fixating the faces, this would confound the perceptual factors of figure-ground with fixation patterns. Therefore we included the following task to aid subjects in maintaining their fixation. A square containing an X was positioned at the blind spot so that accurate fixation would render it invisible (*5*). A trial was initiated only if the blind-spot stimulus was not visible. This precaution eliminated as far as possible the effects of eye movements and fixation location on the performance of the perceptual task (*6*).

On each trial, the observer fixated the point in the center of the display, noted that the blind-spot stimulus was invisible, and pressed a key to present the target line. On one block of trials, the observers initiated the trial only when they perceived the central region as the vase and on another block of trials only when they saw the flanking regions as two faces. Because the target appeared randomly within the central and flanking regions from trial to trial, the contextual effect of each region as figure or as ground could be evaluated. We also included a block of trials in which the target line was presented alone in a homogeneous dark field. A two-alternative forced-choice method was used, and the data were analyzed according to signal detection theory (*7*).

In condition 1, the luminance of the display was set at threshold (75 percent correct), while the tilt angle of the target was set above its discrimination threshold (80 percent correct) (*8*). A tilt angle of 1.6° gave this desired accuracy. In condition 2, the target's tilt angle was set at discrimination threshold (0.8° tilt from the vertical), while the display luminance was set above threshold (*9*).

The results are shown in Fig. 2. Table 1 shows the mean d' (discrimination index) across five observers from the two conditions. Observers performed significantly better when the region surrounding the target was perceived as figure than when it was perceived as ground even though the physical stimulus remained unchanged. The d' for the target presented against the ground was actually lower than that for targets presented alone in the visual field. Thus, the better discrimination of the target within the figure cannot be due to luminance summation of target and context (*10*).

Discrimination for targets flashed at fixation was somewhat better than for targets flashed in locations 1° left or right of fixation. This result is consistent with data showing that visual resolution decreases with distance from the fovea (*11*). However, this decrease of accuracy was constant across all stimulus conditions and did not interactively affect a particular condition (Table 1) (*12*).

Our findings reveal that discrimination can be affected by whether a context is perceived as figure or as ground as well as by the more elementary factors such as luminance and receptive field characteristics. Our findings also add to the growing class of context effects showing that perceptual variables, specifically figure, improve orientation discrimination of a line segment (*13*). Our results demonstrate that such perceptual context effects can influence orientation discrimination accuracy even when a physical stimulus stays the same.

Perception of figure and ground has been suggested to involve two systems with different information processing characteristics (*14*). Figure perception is characterized by detail analysis and high resolution, while ground perception is characterized by low resolution and insensitivity to phase information. Our findings support this dichotomy of figure- versus ground-analysis as basic descriptors of visual processing in addition to the more established dichotomies of sustained versus transient channels and central versus peripheral vision.

Note added in proof: It has been brought to our attention that an earlier experiment (*15*) apparently found the opposite effect on figure-ground thresholds using different stimuli and procedures.

EVA WONG
NAOMI WEISSTEIN
Department of Psychology, State University of New York, Buffalo 14226

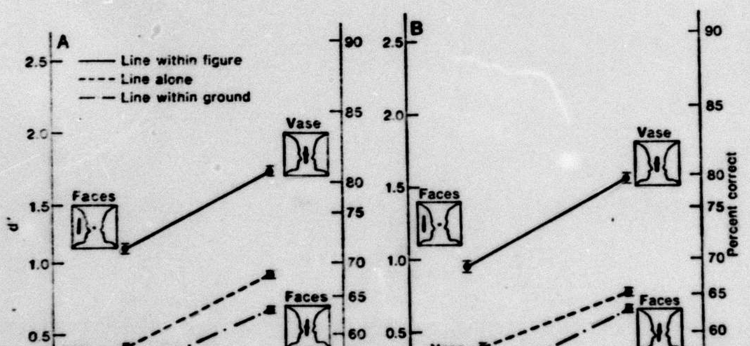


Fig. 2. (A) Display luminance was set at threshold while the tilt angle of the target was set above its discrimination threshold. (B) Tilt angle of the target was set above discrimination threshold while the luminance of the display was set at its discrimination threshold.

References and Notes

1. N. Weisstein and C. S. Harris, *Science* **186**, 752 (1974); N. Weisstein *et al.*, *Perception*, in press; M. Lanze, N. Weisstein, J. R. Harris, *Percept. Psychophys.* **31**, 376 (1982). Metacontrast masking functions, flicker frequency functions, and visual persistence durations can also be affected by such factors as the perceived depth and connectedness of a pattern [M. C. Williams and N. Weisstein, *Vision Res.* **21**, 631 (1981); R. Genter and N. Weisstein, in preparation; N. Weisstein and C. S. Harris, in *Visual Coding and Adaptability*, C. S. Harris, Ed. (Erlbaum, Hillsdale, N.J., 1981), pp. 317–364].
2. J. Baron, in *Handbook of Learning and Cognitive Processes*, W. K. Estes, Ed. (Erlbaum, Hillsdale, N.J., 1978), vol. 6, pp. 131–166.
3. I. Biederman, *Science* **177**, 77 (1972).
4. E. Rubin, in *Readings in Perception*, D. C. Beardslee and M. Wertheimer, Eds. (Van Nostrand, Princeton, N.J., 1958), pp. 194–203.
5. A. Stoper, *Percept. Psychophys.* **13**, 210 (1973).
6. Because it is still possible for observers to shift their gaze just before they initiated a trial, we ran an auxiliary experiment to monitor their fixation patterns during figure-ground reversals. In this experiment, observers detected a pattern flashed at the region of the blind spot when they perceived faces in the Rubin picture and when they perceived a vase while fixating the fixation stimulus. On any trial, the pattern had a .5 probability of being presented. If observers' fixations deviated by more than 0.5° from the fixation point, detection of the blind-spot stimulus should be above chance level. Observers' accuracy in both conditions was ≤ .56. Thus, observers did not significantly shift their gaze away from the fixation stimulus in the center of the Rubin picture by more than 0.5° when they perceived the faces. This control procedure monitors eye position with an accuracy of approximately 0.5° [E. Hering, *Spatial Sense and Movements of the Eye* (American Academy of Optometry, Baltimore, 1942)]. A further control for eye position was the presentation of targets randomly at each of three locations.
7. D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics* (Wiley, New York, 1966). Data from discrimination tasks conducted according to a two-alternative forced-choice method can be analyzed in the same fashion as that from a yes-no detection task. Responding "left" when the target was tilted left constitutes a hit and responding "left" when the target was tilted right constitutes a false alarm. In our experiments, the estimated d' was based on the receiver operating characteristic (ROC) line drawn through the yes-no point. Data from two experienced psychophysical observers included confidence ratings on a six-point scale. These data best fitted ROC lines with unit slopes passing through the yes-no point, suggesting that the ROC curves did not vary as a function of d'.
8. Two preliminary experiments were run to estimate the luminance and the minimum tilt of the target required for 75 percent correct in discriminating whether the target was tilted left or right. The resulting display luminance ranged from 3.1 cd/m² to 3.9 cd/m² across five observers; for all observers, a tilt of 0.8° placed tilt discrimination at 70 percent correct. The parameter estimation by sequential testing (PEST) procedure [M. M. Taylor and C. D. Creelman, *J. Acoust. Soc. Am.* **41**, 782 (1977)] was used to determine the tilt threshold. Luminance thresholds were obtained for two tilt angles—0.8° and 1.6° from the vertical—by adjusting the luminance of the display until the accuracy was 75 percent correct.
9. Since the luminance threshold was more variable between individual observers than the tilt threshold, the luminance of the display was adjusted according to each observer's threshold in both conditions.
10. The randomizing of target location is perhaps the best argument against an eye movement explanation of our results. According to an eye movement account, observers would tend to move their eyes towards the flanking areas when these were perceived as faces. For example, if observers perceived faces, they might move their eyes to the left. Fixating the middle of the left face would improve discrimination if the target appeared there. But the target appeared there randomly only one-third of the time. The rest of the time it appeared at the center or at the right. In this case, looking at the left face would decrease discrimination because the target would be viewed off fixation. Overall performance would suffer if fixation were directed at any location other than the center. Hence, if observers moved their eyes to the flanking regions when these were perceived as figure, they should be less accurate than if they maintained fixation. But observers were more accurate when the flanking regions were perceived as figure than when they were perceived as ground. Therefore eye movements cannot explain these increases in accuracy.
11. S. M. Anstis, *Vision Res.* **14**, 589 (1974); C. A. Johnson, J. L. Keltner, F. Balestrery, *ibid.* **18**, 1217 (1978); R. J. Jacobs, *ibid.* **19**, 1187 (1979); I. Lie, *ibid.* **20**, 967 (1980).
12. This implies that receptive field size and cortical magnification as one moves away from the fovea, as well as perceptual factors of figure and ground, contribute to discrimination performance.
13. Recent evidence shows that a figure in an Escher reversible picture is more likely to be seen to displace during an eye movement than a ground is (B. Bridgeman, *Acta Psychol.*, in press).
14. B. Julesz, in *Formal Theories of Perception*, E. Leeuwenberg and H. Buffart, Eds. (Wiley, New York, 1978), pp. 205–216.
15. A. Gelb and R. Granit, cited by K. Koffka, *Principles of Gestalt Psychology* (Harcourt Brace, New York, 1935).

16. Supported by contract MDA-903-80-C-0202 from the Defense Advanced Research Projects Agency and by grant EY-03432 from the National Eye Institute.

8 June 1981; revised 4 January 1982

Revised

Sharp Targets are Detected Better Against a Figure and Blurred Targets are Detected Better Against a Background[1]

Eva Wong and Naomi Weisstein
State University of New York at Buffalo

Running Head: Figure-ground and spatial frequency

Abstract

There is growing evidence that the performance of perceptual tasks is often facilitated by perceived figure-ness. Accuracy in detection and discrimination of targets is higher when the targets are presented in figural regions than when they are presented in ground regions of an image. This "figure-superiority" might be a result of a functional specialization in the visual analysis of figure; recent theories have also assumed a functional specialization in the visual analysis of ground. If so, we might expect "ground superiority" in situations where task performance requires information available primarily through analysis of ground. We manipulated the spatial frequency of a small line segment and found that when it was sharp (the high spatial frequency components were present) it was detected better in *figural* regions but when we blurred it (only the low to medium spatial frequencies were present) it was detected better in ground regions. These findings support the view that figure and ground analyses involve different specialized functions.

Figure has been demonstrated to facilitate the performance of perceptual tasks in a number of studies. For example, contour discontinuity is better detected in an area perceived as figure than in an area perceived as ground (Weitzman, 1963); retinal image displacement is more visible in a figural region than in a ground region of an Escher picture (Bridgeman, 1981). Recently we have found that the orientation of a tilted line is discriminated more accurately when it is flashed in the figural region than in the ground region of Rubin's reversible goblet-faces picture (Wong & Weisstein, 1982).[2]

The perceptual advantage of figure over ground has been conceptualized in a number of ways. It has been proposed that more attention is given to figure while details of the background are generally ignored. Therefore, motion, displacement, and contour are more salient in a figure than in ground (Bridgeman, 1981). The Gestalt theorists described the figure as having a "thing-like" character (Rubin, 1921) and being "more strongly structured, and more impressive" (Koffka, 1935). Rather than emphasizing the consistent dominance of figure over ground, however, they chose to view the dichotomy of figure and ground in terms of a functional difference. This approach to figure and ground was not further explored theoretically until quite recently, when Julesz proposed that not only different visual processes mediate the processing of figure and ground, but that figure analysis and ground analysis each involve highly specialized functions (Julesz, 1978). Figure analysis is concerned with high resolution of details while ground analysis is concerned with extraction of global background information. Such an approach to figure and ground perception leads to the prediction that in certain situations, ground might facilitate perception while figure would not. This would follow if perceptual performance is dependent on the kind of task as well as the type of information needed to perform the task.

What type of tasks might favor a figure analysis and which types might favor ground analysis? Here we might consider channels in the visual system sensitive to different ranges of spatial and temporal frequencies (Campbell & Robson, 1968; Robson, 1966). These have been hypothesized to have different perceptual functions (Weisstein, 1968; Kulikowski & Tolhurst, 1973; Weisstein, Ozog & Szoc, 1975; Breitmeyer & Ganz, 1976; Weisstein & Harris, 1980). A process sensitive to high spatial frequencies--edges, detail, and phase--might be a good candidate for figure analysis if it is assumed (Julesz, 1978) that such analysis is concerned with detail resolution and vernier signal processing. A process sensitive to low spatial frequencies might be better suited to a ground analysis if ground is assumed concerned with the extraction of global information. It is also assumed that the extraction of global information proceeds more quickly than the extraction of details about a pattern, and some investigators characterize ground analysis as serving an "early warning system" for visual processing (Breitmeyer & Ganz, 1976; Julesz, 1978; Calis & Leeuwenberg, 1981). This differential speed of ground processing is somewhat suggested by the evidence that visual response is <u>faster</u> to low spatial frequencies (Breitmeyer, 1975; Tolhurst, 1975; Vassilev & Mitov, 1976; Breitmeyer & Ganz, 1975; Watson & Nachimas, 1977).

If ground analysis involves channels sensitive to low spatial (and high temporal) frequencies and figure analysis involves channels sensitive to high spatial (and low temporal) frequencies then it is possible that, depending on what spatial frequencies are present in a particular target, one or the other process might predominate and differentially aid perceptual performance.

At threshold the most sensitive system mediates the detection of the stimulus. A theory which identifies different spatial and temporal frequency channels with different perceptual functions would predict that the targets with high spatial frequency components will be detected better in a figural region but targets with only low spatial frequencies present will be detected better in a ground region.

While figure has been shown to improve performance in a variety of tasks, there is as yet no direct evidence showing differential dominance of figure and ground under different stimulus and task situations. The experiments reported here were designed to test whether such differential dominance could be found, using targets of different spatial frequencies.

We ran three experiments. The first was designed to select observers who could hold their gaze at fixation. The second established a luminance level at which the sharp and blurred targets were detected correctly 70% of the time. The third experiment measured the detection of sharp and blurred targets when they were flashed against a figure or against a ground region.

In our experiments, we chose Rubin's reversible goblet-faces picture as our stimulus (see Figure 1). In this display, the physical aspects of

---

Insert Figure 1 about here

---

the stimulus are identical whether a region is seen as figure or as ground. Therefore, the purely perceptual effect of figure and ground can be examined while the local and global environments, spatial frequency and phase, are kept invariant. The target was a vertical line segment presented at threshold at one of three possible positions in the Rubin picture: the central region, the right, or the left flanking regions (Locations A, B, C in Figure 1). In one session, observers initiated a trial only if they perceived the central region as a goblet (or figure) and in another session, they initiated a trial only if they perceived faces (that is, if the flanking regions were seen as figures). Since the target appeared randomly at one of three regions from trial to trial, the effect of perceived figure and ground on target detection could be evaluated for all three perceptual areas.

EXPERIMENT 1

In our experiments it is critical that the observer fixate the fixation stimulus accurately throughout the figure-ground reversals. If observers shifted fixation to the flanking region when it was perceived as figure, and if on that particular trial the target was flashed in the central region of the display, the target would be viewed off-fixation. Consequently, figure-ground effects would be confounded with fixation patterns. Experiment 1 was conducted to select observers for participation in the subsequent experiments who could maintain a given fixation while monitoring figure-ground reversals.

The observer's task was to detect the presence of a stimulus pattern flashed in the blind-spot region of the visual field while maintaining fixation on the fixation stimulus in the center of the display. There were two conditions: in one condition, observers initiated a trial only if they perceived a goblet, and in the other condition they initiated a trial only if they saw two faces. The target pattern had a probability of .5 of being presented, so that if the observer were fixating the fixation point during the trial, detection performance should not deviate significantly from chance level. If the observer's fixation is generally unsteady, then detection performance would be above chance level whether the central region of the Rubin picture was perceived as figure or as a background. If the observer had a tendency to look at figural regions, then detection performance would be above chance level in the condition when two faces were perceived. This procedure of monitoring fixation is accurate to within 1/2 degree (Hering, 1942).

Method

Subjects. Subjects were drawn from an undergraduate subject pool. Two paid experienced psychophysical observers also participated. All had normal or corrected-to-normal vision.

Stimuli. The stimulus consisted of the outline of Rubin's reversible goblet-faces picture covering an area of 5.4 degrees by 5.4 degrees with a dim fixation point in the center of the display.[3] The target pattern presented in the blind-spot region of the visual field consisted of a square enclosing an 'x'. The stimuli were drawn on a CRT screen by a GT-40 graphics display processor controlled by a PDP-11 computer.

Procedure. At the beginning of each trial, the observer fixated the dim fixation point in the center of the display. Viewing was monocular; the right eye was occluded. The "blind-spot" target was turned on and its position adjusted so that accurate fixation of the fixation stimulus would render this pattern invisible. The blind-spot pattern was then extinguished. In the first block of trials, observers pressed a key to initiate a trial when they perceived the central region as a goblet. In the second block of trials, they initiated a trial only when they perceived the flanking regions as faces. On any given trial the blind-spot target was flashed for 20 msec. and was presented with a probability of .5. In each block there were a total of 75 signal trials and 75 trials where no target was presented. Observers made a forced-choice "yes-no" response.

Results

Five observers were selected out of twelve candidates tested from the undergraduate subject pool. These observers' detection accuracy as well as that of the two experienced observers did not exceed a probability of .6. The mean probabilities for the seven observers for detection of the blind spot stimulus were .54 (S·D = .03) when the central region was perceived as figure and .55 (S.D = .04) when the flanking regions were perceived as figure.

EXPERIMENT 2

Before we can compare the detectability of targets of different spatial frequencies at threshold in figure and in ground regions, it is necessary to find luminance levels where both the sharp and blurred targets are detected correctly with a probability of .7. This experiment was run to obtain such a threshold for each subject.

Method

Subjects. The subjects were those selected from Experiment 1.

Stimuli. Since the threshold of a briefly flashed vertical line segment depends on what is also present in the visual field, the display consisted of a "neutral" unambiguous figure whose line length matched that of the Rubin picture (See Figure 2).

Insert Figure 2 about here

The "neutral" figure subtended an area of 5.4 degrees by 5.4 degrees square. It was divided into three regions, the areas approximating the three regions in the Rubin picture. The unblurred target was a vertical line .9 degree in length and .06 degree wide. A dim fixation point was located in the center of the display; the target was presented in this position.

The spatial frequency of the target was manipulated by blurring (Gonzalez & Wintz, 1978; Zucker, 1980). Blurring was first approximated by placing frosted acetate over the screen. The resulting two-dimensional intensity profile was charted using a photometer (Gamma Scientific Model 2900). The desired intensity profile was simulated by computer using neighborhood spatial averaging techniques (See Appendix for details).[4] This profile was then displayed on the CRT of the GT-40 graphics display processor.

Figure 3 shows the intensity profiles of the sharp and blurred targets, and their respective spatial frequency spectra. The spatial frequency spectrum of the sharp target is relatively low and flat, with energy at very high frequencies, while that of the blurred target starts high and falls off quite rapidly with increasing frequencies. Thus, at a frequency of 8.3 c/deg (the "fundamental" frequency, that is, the reciprocal of twice the width of the sharp target--see Weisstein, 1980, for an introductory discussion of Fourier analysis) the blurred target has a higher amplitude than the sharp target, while beyond 15.6 c/deg the amplitude of the blurred target is essentially at zero.

Insert Figure 3 about here

Procedure. There were two stimulus conditions which were blocked into two sessions. In one session, the target was a sharp line and in the other, the target was blurred. For our exposure duration (20 msec. flashes) we would not expect any significant threshold difference between blurred and sharp targets (Hood, 1973) but we obtained measurements to ensure that our subjects did not deviate from this norm.

Each trial began with the observer fixating the dim fixation spot in the center of the display. The subject pressed a key to initiate a trial and the target was flashed for 20 msec. at the location of the fixation point. The fixation stimulus was extinguished just before the target was presented and reappeared after the target blanked. On each trial the target had a .5 probability of being presented. On trials where no target was presented the fixation stimulus was blanked for 20 msec. The observer made a "yes-no" response. After a block of 50 signal trials and 50 non-signal trials, the percentage of correct responses was calculated and the luminance of the target adjusted until the subject obtained an accuracy of 70% for three consecutive blocks of trials. The luminance setting for each particular observer was subsequently used in Experiment 3. The detection threshold of the sharp target was determined first. In the second session, the luminance of the blurred target was initially set to match that obtained for the sharp target and adjusted if necessary.

Results

The aperture of the photometer was set so that it measured total luminous flux in both the horizontal and vertical directions for both blurred and sharp targets. With total luminous flux constant, we found no significant threshold difference between the sharp and blurred targets for any of the seven subjects. Therefore the luminance of both targets was matched in the third experiment. This meant (see Figure 3) that not only were the high spatial frequencies absent in the blurred target, but also the energy in the low end of the spectrum was greater.

EXPERIMENT 3

In this experiment the detection of sharp and blurred targets flashed against figure and ground regions was compared.

Method

Subjects. The subjects were those selected from Experiment 1.

Stimuli. The display was the outline of Rubin's reversible goblet-faces picture (described in Experiment 1). The target was a vertical line .9 degrees in length and .06 degrees wide. Blurring of the target was achieved using the method described in Experiment 2. All the display parameters were the same as those in Experiment 1.

Procedure. The experiment was divided into two sessions, run on two different days. In the first session the target was a sharp line segment; in the second session the target was blurred. On a given trial, the target was presented randomly at one of three locations in the Rubin picture (locations A, B, C in Figure 1).

There were two blocks of trials in a session. In the first block, observers initiated a trial only if they perceived a goblet (that is if the central region was seen as figure) and in the second block, they initiated the trial only if two faces were perceived (that is, if the flanking regions were seen as figure). The observer always fixated the dim fixation spot in the center of the display. Since on any given trial the target randomly appeared in one of the three regions, four viewing conditions were generated with the target viewed, respectively, at fixation in a figural region; off fixation in a figural region; at fixation in a ground region; off-fixation in a ground region.

At the beginning of a session, the blind-spot pattern (See Experiment 1) was turned on and its position adjusted so that accurate fixation on the fixation stimulus made this pattern invisible. Viewing was monocular; the right eye was occluded. The observer fixated the fixation stimulus, noted that the blind-spot pattern was invisible, and initiated a trial by pressing a key. On any trial the target had a probability of .5 of being presented in the central region and .25 of being presented in each of the flanking regions. There were a total of 200 signal trials where the target was presented in the central area; 200 trials where the target was presented in the flanking areas; and 200 non-signal trials. The luminance of the target was set at the particular observer's threshold obtained from Experiment 2. The target was flashed for 20 msec; the fixation stimulus was extinguished just before the target was presented and reappeared after the target blanked. In non-signal trials the fixation stimulus also blanked for a period of 20 msec. The observer made a forced-choice "yes-no" response and rated his or her confidence on a six-point scale.

Results

The data were analyzed by means of signal detection theory. In both single-level binary decision and confidence rating procedures, the probabilities p(s/sn) (responding "yes" given a signal) and p(s/n) (responding "yes" given no signal) are transformed so that the normalized z scores are linearly spaced along the coordinate axes of p(s/sn) and p(s/n). Data generated by these procedures are best fitted by an ROC (Receiver Operating Characteristic) line with unity slope. However, it is possible for the true ROC curve to have a non-unity slope.

We found each subject's slope to be slightly less than unity, (it is common in visual psychophysics for the ROC curve to be somewhat less than 1). This was taken into account by using the parameter d_s to estimate d'. In a plot of normal-normal co-ordinates with p(s/n) and p(s/sn) occupying the x and y axes respectively, d_s can be read off directly from the point of intersection between the operating characteristic and the negative diagonal (Clarke, Birdsall, & Tanner, 1959; Egan, Schulman & Greenberg, 1959). The d's for each subject were estimated using d_s.

Figure 4 summarizes the data across all the observers from this experiment.

Insert Figure 4 about here

When sharp lines were presented in figural regions, the signal-to-noise ratio was better than when they were flashed in a ground region. However, when the target line was blurred the signal-to-noise ratio was better when the targets were flashed in the ground region. This result occurred in both on-fixation and off-fixation viewing conditions. Thus, sharp targets--those with high spatial frequencies present--are detected better against figure, while blurred targets--those with lower spatial frequencies--are detected better against ground. Off-fixation viewing attenuated the d' across both figure and ground contexts by a constant magnitude. This indicates that an early processing constraint of retinal eccentricity on detection was present in addition to the perceptual effects of figure and ground and is consistent with the fall-off in resolution as a function of distance from the fovea (Anstis, 1974; Johnson, Keltner, & Balestrery, 1979; Jacobs, 1979; Lie, 1980).

A three-way ANOVA (repeated measures: fixation; spatial frequency; figure or ground) showed the main effect of fixation to be significant ($F = 6.32$; $df = 1,6$; $p < .05$). The interaction of figure/ground and target spatial frequency is highly significant ($F = 20.18$, $df = 1,6$; $p < .005$). T-tests were performed between the cell means which involved the interaction of figure-ground and sharp and blurred targets. T values which are significant at the .005 level (two tailed) were obtained for the following four comparisons: 1) Figure/at fixation/sharp target vs. Ground/at fixation/sharp target, 2) Figure/off fixation/sharp target vs. Ground/off fixation/sharp target, 3) Figure/at fixation/blurred target vs. Ground/at fixation/blurred target, 4) Figure/off fixation/blurred target vs. Ground/off-fixation/blurred target. These results indicate that there is a marked superiority of figure over ground for sharp targets and a marked superiority of ground over figure for blurred targets. None of the other interactions are significant.

Discussion

Targets with high spatial frequency components were detected better when they were flashed in figural regions than when they were flashed in ground regions. Blurred targets, on the other hand, were detected better when they were flashed in ground regions than when they were flashed in figure regions. This finding is in line with the theory that different visual processes mediate the analysis of figure and ground.

It has been proposed that an early, global extraction stage of visual processing is passive, pre-attentive and effortless while a later stage, where details are scrutinized requires active attention (Broadbent, 1977; see also Navon, 1981; Miller, 1981a, b, for an extended discussion). From our findings it is not possible to conclude that figure analysis involves effortful scrutiny of details and ground analysis is a passive extraction of information. Rather, if effortful scrutiny or attention is indicated by better detection performance, than such scrutiny would seem to have little to do with our results. The region of the display where it would be plausible for observers to direct their attention--the figure region--did not always yield the best detection accuracy. Moreover, while fixation is not always a firm measure of attention (see Posner, Snyder & Davidson, 1980 for a discussion of focal attention decoupled from direction of gaze) it is worth nothing that the selective facilitation of either figure or ground occurred in off-fixation as well as on-fixation conditions. What determined relative accuracy in our experiments does not appear tied to attention; instead, it involves the interaction of figure and ground with different spatial frequency ranges.

Channels sensitive to low spatial frequencies have been thought to be involved in extraction of background global information (Henning, Hertz, & Broadbent, 1975; Broadbent, 1977). In other words, ground analysis has been thought to involve the low end of the frequency spectrum. Moreover, if ground analysis were to serve as an "early warning system", information must be available earlier for decision and action. There is much psychophysical evidence that channels tuned to low spatial frequencies have a faster response than those tuned to high spatial frequencies: they appear to have a shorter latency (Breitmeyer, 1975; Vassilev & Mitov, 1975; Lupp, Hauske, & Wolf, 1976), a faster rise-time (Robson, 1955; Wantanabe, Mori, Nagata, & Hiwatashi, 1976; Breitmeyer & Ganz, 1976; Watson & Nachmias, 1977; van Nes, Koenderink, & Bouman, 1978) and shorter integration time constants (Nachmias, 1967; Tolhurst, 1975). Thus channels sensitive to low spatial frequencies would seem best suited for processing ground in two ways: they respond fast, and they respond to more global features of a stimulus.

Although it has been previously suggested that low spatial frequency channels are involved in ground analysis, (and high spatial frequency channels are involved in figure analysis), direct evidence for this has been lacking up until now. Our findings provide such evidence, linking high and low spatial frequencies to figure and ground processing, respectively.

Calis and Leeuwenberg (1980) have also argued that ground analysis occurs earlier in visual processing than figure analysis. They used a "probing" technique which distorted the first of a pair of stimuli by presenting a second one at varying stimulus onset asynchronies (SOAs). When the first stimulus formed the ground of an object, maximal distortion of ground occurred at shorter SOAs than when the first stimulus formed a figure. This suggests that ground analysis must be occurring earlier than figure analysis. The implication of our findings that sharp targets are enhanced in figural regions while blurred targets are enhanced in ground regions are closely related to those of Calis and Leeuwenberg's although their interest is in the formation of the percept in real time while here we are more concerned with the kind of information involved in figure and ground analyses. They have charted the process of percept formation and showed that information about ground is extracted before information about figure. Our experiments show that low spatial frequencies as well are linked to ground analysis, thus giving a basis as to how the information used in the "coding" of ground might be made available earlier to post-sensory processing.

Appendix

Method of blurring the target

Given an N by N image $f(x,y)$, the aim is to generate an image $g(x,y)$ whose intensity at every point (x,y) is obtained by averaging the intensity values of the display screen matrix f contained in a defined neighborhood of (x,y). The procedure is defined by the equation $g(x,y) = \frac{1}{M} \sum_{(m,n)\ S} f(m,n)$ for $x,y = 0, 1, 2\ldots\ldots N-1$. S is the set of coordinates of points of the display matrix in the neighborhood of point (x,y). M is the total number of points in set S. The boundary of the set S is determined by the radius of a circle centering on point (x,y). The degree of blurring is highly correlated with the radius. By manipulating the radius size, a two-dimensional intensity profile which will approximate the blurring may be obtained. A neighborhood radius equal to 8 was used to blurr the target used in Experiments 2 and 3.

References

Anstis, S. M. A chart demonstrating variations in acuity with retinal position. Vision Research, 1974, 14, 589-592.

Breitmeyer, B. G. Simple reaction time as a measure of the temporal response properties of transient and sustained channels. Vision Research, 1975, 15, 1411-1412.

Breitmeyer, B. G., & Ganz, L. Implications of sustained and transient channels for theories of visual pattern masking, saccadic suppression, and information processing. Psychological Review, 1976, 83, 1-36.

Bridgeman, B. Cognitive factors in subjective stabilization of the visual world. Acta Psychologica, 1981, 48, 111-121.

Broadbent, D. E. The hidden preattentive process. American Psychologist, 1977, 32, 109-118.

Calis, G., & Leeuwenberg, E. Grounding the figure. Journal of Experimental Psychology: Human Perception and Performance, 1981, 7, 1386-1397.

Campbell, F. W., & Robson, J. G. (1968) Application of Fourier analysis to the visibility of gratings. Journal of Physiology, 197, 551-566.

Clarke, F. R., Birdsall, T. G., & Tanner, W. P. Jr. Two types of ROC curves and definitions of parameters. Journal of the Acoustical Society of America, 1959, 31, 629-630.

Egan, J. P., Schulman, A. I., & Greenberg, G. Operating characteristics determined by binary decisions and by ratings. Journal of the Acoustical Society of America, 1959, 31, 768-773.

Gelb, A., & Granit, R. Die Bedeutung von "Figur" und "Grund" fur die Farbenschwelle. Zts. f. Psych., 1923, 93, 83-118.

Gonzalez, R. C., & Wintz, P. Digital image processing. Reading: Addison-Wesley, 1977.

Green, D. M., & Swets, J. A. Signal detection theory and psychophysics. New York: Wiley, 1966.

Henning, G. B., Hertz, B. G., & Broadbent, D. E. Some experiments bearing on the hypothesis that the visual system analyses spatial patterns in independent bands of spatial frequency. Vision Research, 1975, 15, 887-898.

Hering, E. Spatial sense and movements of the eye, A. Radde, trans., American Academy of Optometry, 1942.

Hood, D. C. The effects of edge sharpness and exposure duration on detection threshold. Vision Research, 1973, 13, 759-766.

Jacobs, R. J. Visual resolution and contour interaction in the fovea and periphery. Vision Research, 1979, 19, 1187-1196.

Johnson, C. A., Keltner, J. L., & Balestrery, F. Effects of target size and eccentricity on visual detection and resolution. Vision Research, 1978, 18, 1217-1222.

Julesz, B. Perceptual limits of texture discrimination and their implications to figure-ground separation, in E. Leeuwenberg (Ed.), Formal theories of perception, New York: Wiley, 1978.

Koffka, K. Principles of Gestalt psychology. New York: Harcourt-Brace, 1935.

Kulikowski, J. J., & Tolhurst, D. J. Psychophysical evidence for sustained and transient detectors in human vision. Journal of Physiology, 1973, 232, 777-784.

Lie, I. Visual detection and resolution as a function of retinal locus. Vision Research, 1980, 20, 967-974.

Lupp, U., Hauske, G., & Wolf, W. Perceptual latencies to sinusoidal gratings. Vision Research, 1976, 16, 969-972.

Miller, J. Global precedence in attention and decision. Journal of Experimental Psychology: Human Perception & Performance, 1981, 7, 1161-1174.

Miller, J. Global precedence; information availability or use? Reply to Navon. Journal of Experimental Psychology: Human Perception & Performance, 1981, 7, 1183-1185.

Nachmias, J. Effect of exposure duration on visual contrast sensitivity with square-wave gratings. Journal of the Optical Society of America, 1967, 57, 421-427.

Navon, D. Forest before trees: The precedence of global features in visual perception. Cognitive Psychology, 1977, 9, 353-383.

Navon, D. Do attention and decision follow perception? Comment on Miller. Journal of Experimental Psychology: Human Perception & Performance, 1981, 7, 1175-1182.

Osgood, C. E. Method and theory in experimental psychology, New York: Oxford University Press, 1953.

Posner, M. I., Synder, C. R. R., & Davidson, B. J. Attention and the detection of signals. Journal of Experimental Psychology: General, 1980, 109, 160-174.

Robson, J. G. Spatial and temporal contrast sensitivity functions of the visual system. Journal of the Optical Society of America, 1966, 56, 1441-1442.

Rubin, E. Figure and ground (1921), In D. C. Beardslee and M. Wertheimer (Eds.), Readings in perception. Princeton: Van Norstrand, 1958.

Tolhurst, D. J. Sustained and transient channels in human vision. Vision Research, 1975, 15, 1151-1155.

Vassilev, A., & Mitov, D. Perception time and spatial frequency. Vision Research, 1976, 16, 89-92.

Van Nes, F. L., Koenderink, J. J., & Bouman, M. A. Spatio-temporal modulation transfer in the human eye. Journal of the Optical Society of America, 1967, 57, 1082-1087.

Wantanabe, A., Mori, T., Nagata, S., & Hiwatashi, K. Spatial sine-wave responses of the human visual system. Vision Research, 1968, 8, 1245-1263.

Watson, A. B., & Nachmias, J. Patterns of temporal interaction in the detection of gratings. Vision Research, 1977, 17, 893-902.

Weisstein, N. The joy of Fourier analysis. In C. S. Harris (Ed.) Visual coding and adaptability. Hillsdale, N. J.: Erlbaum, 1980.

Weisstein, N. A Rashevsky-Landahl neural net: Simulation of metacontrast. Psychological Review, 1968, 75, 494-521.

Weisstein, N., Ozog, G., & Szoc, R. A. A comparison and elaboration of two models of metacontrast. Psychological Review, 1975, 82, 375-343.

Weisstein, N., & Harris, C. S. Masking and the unmasking of distributed representations in the visual system. In C. S. Harris (Ed.), Visual coding and adaptability. Hillsdale, N. J.: Erlbaum, 1980.

Weitzman, B. A threshold difference produced by a figure-ground dichotomy. Journal of Experimental Psychology, 1963, 66, 201-205.

Wong, E., & Weisstein, N. A new perceptual context-superiority effect: Line segments are more visible against a figure than against a ground. Science, 1982, 218, 587-589.

Zucker, S. W. Motion and the Mueller-Lyer illusion. Technical Report, McGill University, Department of Engineering, 1980.

Footnotes

[1]This work was supported by grant R01 EY03432 from the National Eye Institute to the second author. We thank C. S. Harris for his helpful comments.

[2]But see Koffka (1935) who cited Granit and Gelb (1923) as apparently having found the opposite effect: Spots of light were detected at lower thresholds on ground than on figure regions. Osgood (1953) provides a possible explanation for this seeming "figure inferiority" in terms of edge effects.

[3]In an earlier study (Wong & Weisstein, 1982) the Rubin figure subtended 3.2 by 3.2 degrees. Our figure-ground effects are therefore not confined to one stimulus dimension.

[4]Blurring in the spatial domain is essentially equivalent to lowpass filtering in the frequency domain. In spatial domains, this can be described by the relation.

$$g(x,y) = h(x,y)*f(x,y) \tag{1}$$

where $g(x,y)$ is the blurred image

$h(x,y)$ is the blurring function

and $f(x,y)$ is the original image

In the frequency domain, this low-pass filtering is described by the relation

$$G(u,v) = H(u,v)F(u,v) \tag{2}$$

where $G(u,v)$ is the Fourier transform of blurred image,

$H(u,v)$ is the Fourier transform of the low-pass filter function

$F(u,v)$ is the Fourier transform of the original image

$h(x,y)$ in the spatial domain can be solved given expression (1). The inverse of h(x,y) yields H(u,v), the low-pass filter function and the associated cut-off frequency. The cutoff frequency for the blurred stimulus as calculated by standard procedures of 2 dimensional Fourier analysis and assuming an ideal low-pass filter is 15.56 cycles/deg.

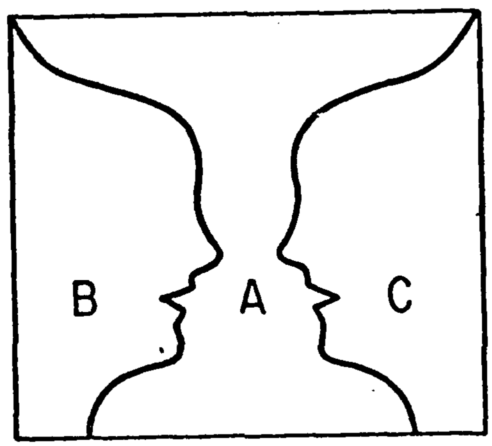
Figure Captions

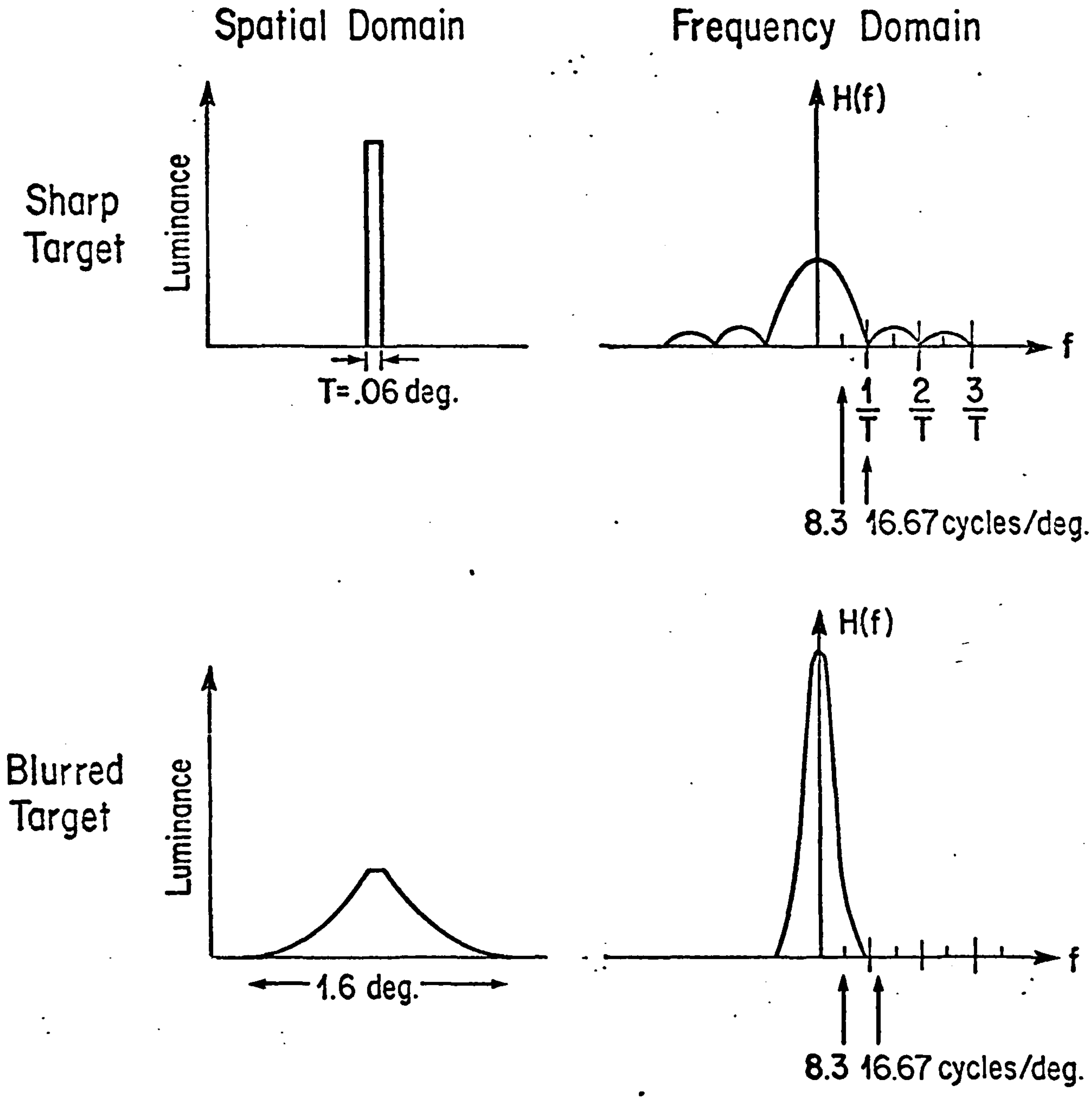
Figure 1. The Rubin goblet-faces picture: Display used in Experiments 1 and 3.

Figure 2. The neutral figure: Display used in Experiment 2.

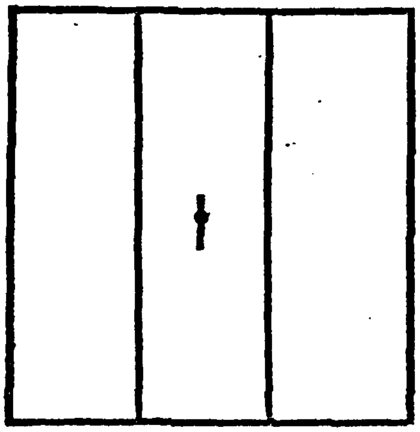
Figure 3. Illustration of the sharp and blurred targets in the spatial domain and the frequency domain. The functions are to scale, except for the y-axis of the frequency domain, where the amplitude of the blurred spectrum at the very low frequencies is about 3 times greater than what is shown. For purposes, of space, this ratio is not kept in the figure.

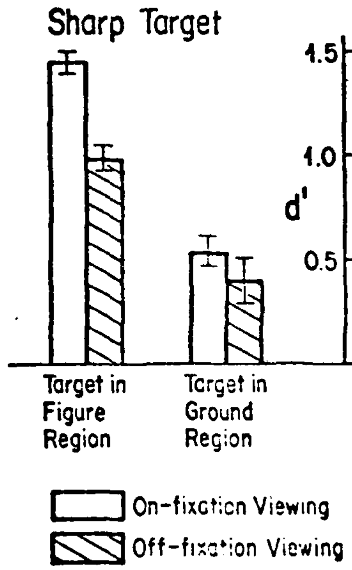
Figure 4. Results of Experiment 3: Detection of sharp and blurred targets against figure and against ground.


B
A
C



Spatial Domain
Frequency Domain
Sharp Target
Luminance
T=.06 deg.
H(f)
f
1/T
2/T
3/T
8.3
16.67 cycles/deg.
Blurred Target
Luminance
1.6 deg.
H(f)
f
8.3
16.67 cycles/deg.




Sharp Target
Blurred Target
1.5
1.0
d'
0.5
Target in Figure Region
Target in Ground Region
Target in Figure Region
Target in Ground Region
On-fixation Viewing
Off-fixation Viewing

Flicker Induces Depth: Spatial and Temporal Factors in the Perceptual Segregation of Flickering and Non-flickering Regions in Depth

Eva Wong and Naomi Weisstein

State University of New York at Buffalo

Running Head: Flicker induces depth

Abstract

If some regions of a random-dot field are flickered, then the non-flickering areas appear to stand out in depth in front of the flickering regions. This perception of depth is optimal within a limited range of temporal frequencies. The average temporal luminance of the flickering and non-flickering regions was kept equal; therefore the depth segregation is not due to a luminance difference. In fact, depth is seen even when the average temporal luminance of the flickering regions is twice that of the steadily-presented one. The magnitude of perceived depth is affected by the percentage of luminance modulation: depth is maximal at 100% modulation and diminishes as the percent modulation decreases. We charted the tuning function using alternating flickering and non-flickering random-dot bars and found it to be similar to those of visual channels most sensitive to high temporal frequency.

Flicker Induces Depth: Spatial and Temporal Factors in the Perceptual Segregation of Flickering and Non-flickering Regions in Depth

Recently we discovered that if areas of a filled visual field are flickered, the flickering regions appear to lie in depth well behind the non-flickering regions (Wong & Weisstein, 1982). When parts of a random-dot field were flickered so that the display consisted of alternating flickering and non-flickering "bars" composed of dots, the non-flickering "bars" were seen to stand out in front of the flickering "bars". This depth segregation produced by flicker--flicker-induced depth--is not due to a luminance difference between the flickering and non-flickering areas since the average temporal luminance of all the regions was kept equal. It is also not dependent on the textural elements making up the visual field since we obtained the effect with dots, horizontal lines and vertical lines; nor does it depend on a specific configuration of the flickering and non-flickering regions since we obtained this effect with "gratings" and with concentric squares. Moreover, we found that a temporal frequency of about 6 Hz. produced the greatest depth separation between the flickering and non-flickering areas, suggesting that visual channels responding primarily to high temporal frequencies might be involved in segregating perceptual regions in depth.

At first glance, this flicker-induced depth may recall the Gestalt organization principle "grouping by common fate" (Wertheimer, 1923; Johansson, 1950). (Johansson has demonstrated that dots moving in the same direction are grouped together.) However, a closer examination reveals that our phenomenon is not just an instance of grouping by flicker, analogous to "grouping by common fate". First, in our case flicker produces depth as well as segregation of the perceptual field into defined regions while "grouping by common fate" typically defined does not involve depth. Second, our phenomenon is optimal around a temporal frequency of 6 Hz., showing an effect of tuning whereas "grouping by common fate", for example grouping by motion, is not dependent on the velocity of the moving elements (Johansson, 1950). Finally, we have some preliminary findings showing that when a region is flickered at very low temporal frequencies (below 2 Hz.), grouping of the flickering and non-flickering elements can still be experienced but the depth separation between those regions has disappeared. Thus although "grouping by flicker" might be the basis of segregating the elements into defined regions, certain rates of temporal modulation are necessary for perceiving the flickering and non-flickering areas as segregated in dept

The three experiments that we report in this paper are further investigations of the depth segregation produced by flicker. The first experiment examined the spatio-temporal tuning of the flicker-induced depth using "gratings" composed of random dots. The second experiment examined the effect of the amplitude of temporal modulation on the amount of perceived depth separation between the flickering and non-flickering areas. In the third experiment, we looked at how luminance differences between the flickering and non-flickering regions might affect the amount of depth perceived.

EXPERIMENT 1

This experiment was designed to investigate the spatio-temporal tuning of depth segregation produced by flicker. .

Method

Subjects. Seven naive observers from an undergraduate subject pool participated in this experiment. All had 20/20 vision.

<u>Stimuli</u> <u>and</u> <u>Apparatus</u>. The display consisted of a field of random dots covering an area of 5.4 degrees by 2.7 degrees (see Figure 1). Regions of this random-dot field were flickered to form a "grating" composed of

---

Insert Figure 1 about here

---

alternating flickering and non-flickering "bars". All "bars" were equal in width. There were nine bar widths (2.7, 1.35, 1.0, .68, .54, .34, .27, .21, .14 degrees) and twelve temporal square-wave frequencies (1, 1.4, 2, 2.8, 3.6, 5, 5.5, 6.3, 7.1, 8.3, 10, 12.5 Hz.).

A PDP-11 computer with a GT-40 graphics processor generated the stimuli and controlled the experiment. The entire random-dot field was composed of two subpictures--the flickering regions and the non-flickering regions. The computer first selected the bar width from a list containing the randomized order of the bar widths to be used in the experiment, calculated the screen coordinates bordering each "bar", and tagged each "bar" as either "flickering" or "non-flickering". The "flickering" and "non-flickering" regions were then filled with random dots to form two respective subpictures, one of the subpictures to be flickered while the other to remain steady on the screen (see Figure 1). The intensity of the dots in the flickering areas was set so that during flicker, the average temporal luminance of the flickering and non-flickering regions was kept equal. (The GT-40 display processor has eight intensity levels and each subpicture can be generated with a specified intensity level.) The mean luminance L_{mean} is defined in terms of $L_{max}+L_{min}$ /2 where L_{max} and L_{min} are respectively the maximum and minimum luminances in the temporal cycle. The temporal modulation amplitude was set at 100% in this experiment and square-wave flicker was used throughout. Flickering was achieved by turning on and off a subpicture. The on/off cycle defined the temporal frequency.

Procedure. Subjects viewed the display from a distance of one metre. Viewing was monocular; the left eye was occluded. At the beginning of each trial the subject fixated a fixation mark (a cross) in the center of the screen. When ready, the subject pressed a key to initiate the trial. The computer selected a certain bar width randomly, generated the display, randomly picked a flicker rate out of the twelve frequencies and presented the flickering stimulus for ten seconds. At the end of each trial the subject first indicated whether flickering and non-flickering dots were segregated into defined regions regardless of any depth separation seen between those areas. A horizontal line then appeared on the screen, and the subject instructed the experimenter to adjust its length so that it matched the amount of perceived depth separation between the flickering and non-flickering areas. The direction was signed (+) if the non-flickering regions were perceived to be in front of the flickering regions. A (-) was signed if the reverse occurred. Each bar width and temporal frequency combination was presented five times. Each data point on the tuning function was thus based on the mean of seven observers each making judgments from five exposures.

Results

First, no observer perceived the flickering areas in front of the non-flickering areas for all spatial and temporal frequency combinations (binomial, $p. < .001$). Second, all observers perceived the flickering and non-flickering dots segregated into definitive perceptual regions regardless of whether depth separation was seen (binomial, $p. < .001$). Third, we were able to obtain a definite tuning function for the depth segregation produced by flicker.

Figure 2 shows the tuning function displayed in a

Insert Figure 2 about here

---

two-dimensional plot. Each point defined by x-y coordinates represents a locus in the tuning function. The iso-depth contours joining the coordinate points connect spatial (bar width) and temporal loci giving a similar level of depth separation perceived between the flickering and non-flickering areas. Each contour (or level of perceived depth) was arrived at by the following way. First, the response measures (adjustment of the length of a horizontal line) from the seven subjects for each spatial and temporal frequency combination were averaged. Thus ninety-six values (from eight bar widths and twelve temporal frequencies) were generated. The computer found the six largest gaps among the ninety-six values. Gap sizes two standard deviates below the mean were dropped; the remaining gaps were then used to define clusters of depth values. Five clusters emerged from the data collected. (See Acher, Shneier, and Rosenfeld (1982) for a discussion and evaluation of this and other methods of cluster extraction.) Within each of the five contours defined by clustering, no subject's judgment fell out of that contour.

From Figure 2 it is clear that large bar widths and high temporal frequencies produced the maximum depth separation between the flickering and non-flickering regions. At temporal frequencies of 6.3 Hz., 7.1 Hz., and bar widths of 1.35 degrees and .68 degrees, the depth effect is greatest. The upper spatial limit (in terms of bar width) is bounded by 2.7 deg. while the depth effect disappears below a temporal frequency of 2 Hz. Depth is perceived up to the highest flicker rate used in this experiment (12.5 Hz.) although it is substantially diminished. At the lower end of smaller bar widths, depth segregation occurs down to .34 deg.

EXPERIMENT 2

This experiment examined the effect of amplitude of temporal modulation on depth segregation between flickering and non-flickering regions. The spatial (bar width) and temporal frequency range which gave the optimal depth response in Experiment 1 was used.

Method

Subjects. Seven new naive observers from the undergraduate subject-pool participated in this experiment. All had 20/20 vision.

Stimuli and Apparatus. The display was the same as that of Experiment 1. Four bar widths of alternating dots (2.7, 1.35, 1.0, .68 degrees) and five temporal frequencies of square wave flicker (5, 5.5, 6.3, 7.1, 8.3 Hz.) were used. These were the spatial and temporal parameters that gave sizable amounts of depth separation between the flickering and non-flickering regions. The flickering regions were modulated at amplitudes of 25%, 50%, 75%, and 100%. Amplitude of modulation is defined as $L_{max} - L_{min} / L_{max} + L_{min}$. At 100% modulation, the stimulus is identical to that in Experiment 1. The average temporal luminance across all spatial regions and at all the modulation amplitudes was kept equal.

The display was generated using the methods described in Experiment 1 except that an additional procedure was needed to simulate the different percentages of temporal modulation. This was achieved in the following way. The non-flickering regions were generated as one subpicture (as in Experiment 1). The flickering regions now consisted of two identical subpictures superimposed on each other. Their intensities were set so that when both of them were turned on, the total luminance equalled the maximum luminance (L_{max}) in the cycle. When one of them was turned on, intensity equalled that of the minimum in the cycle (L_{min}). Thus by changing the intensities of the two subpictures making up the flickering regions, modulation amplitudes of 25%, 50%, 75%, can be simulated. At 100% modulation, one subpicture was set at zero intensity.

Procedure. Viewing conditions were similar to those in Experiment 1. Combinations of bar width and temporal frequency, and modulation amplitude were presented in randomized order. Data were collected in the way described in Experiment 1.

Results

First, no observer perceived the flickering regions in front of the non-flickering regions for all modulation amplitudes across all the bar width and temporal frequencies used in this experiment (binomial, p. < .001).

Second, all subjects perceived the flickering and non-flickering dots as segregrated into definite perceptual regions regardless of whether depth separation was perceived between them (binomial, p. < .001). Third, we found that amplitude of temporal modulation affects the magnitude of perceived depth separation between the flickering and non-flickering areas. In Figure 3, panels A, B, & C, show the spatio-temporal tuning functions of the depth effect for 100%, 75%, and 50% modulation respectively. At 25% modulation, no depth was perceived although the flickering and non-flickering dots were segregated into distinct perceptual regions.

Insert Figure 3 about here

As in Experiment 1, the iso-depth contours here were defined by clusters of values from the response measures. Five clusters were extracted from eighty values (four bar widths, five temporal frequencies, four amplitudes of temporal modulation) each of which was based on the average of seven observers making five adjustments. Within each of the five contours, no subject's depth judgment fell out of that contour.

From Figure 3 it is clear that perceived depth separation between flickering and non-flickering regions was maximal at 100% modulation. Data from this condition follow trends similar to those of Experiment 1 at the same temporal frequencies and bar widths, thus replicating the findings of the previous study with different subjects. At 75% modulation, depth segregation diminished. At the bar widths and temporal frequencies where depth was maximum at 100% modulation (1.35 deg. and .68 deg., 6.3, 7.1, and 8.3 Hz.), the amount of perceived depth has dropped by an average of at least 1 cm. At spatio-temporal regions where depth was marginal at 100% modulation (.34 deg., and below 6.3 Hz.), no depth segregation was perceived at 75%. In terms of contours of depth judgments, the level of depth values has dropped three steps, from the highest level down to the second lowest level. At 50% modulation the amount of depth separation decreased by an average of about 2 cm at bar widths and temporal frequencies where depth was maximum at 100% modulation. In regions where depth was minimal at 75% modulation, depth segregation disappeared at 50% modulation. At 25% modulation no depth was seen at all in the stimulus although segregation of the random dots into flickering and non-flickering regions was perceived. In fact, the striking feature of Figure 3 emerges in the comparison of the three panels (A, B, C): disappearance of the iso-depth contours representing high depth values as depth of modulation decreases, and encroachment of contours representing low depth values into those areas formerly occupied by contours of high depth values.

A 3-way ANOVA (repeated measures) showed all three main effects to be significant (bar width factor: $F = 15.81$, $df = (3,18)$, $p. < .001$; temporal frequency factor: $F = 13.78$, $df = (4,24)$, $p. < .001$; amplitude of modulation: $F = 27.55$, $df = (3,18)$, $p. < .001$). All the two-way interactions were significant (bar width x temporal frequency: $F = 2.54$, $df = (12,72)$, $p. < .01$; bar width x modulation amplitude: $F = 2.07$, $df = (9,54)$, $p. < .05$; temporal frequency x modulation amplitude: $F = 1.96$, $df = (12,72)$, $p. < .05$). The three-way interaction tended toward significance at the .05 level but did not reach the statistical criterion ($F = 0.94$, $df = (36,216)$.

EXPERIMENT 3

In this experiment we examined the effect of luminance differences on the amount of depth perceived between flickering and non-flickering regions. The luminance of the flickering areas was varied while the luminance of the non-flickering areas was kept constant.

Method

Subjects. Seven new naive observers from the undergraduate subject pool participated. All had 20/20 vision.

Stimuli and Apparatus. The display was the same as that used in Experiments 1 and 2. The four bar widths and five flicker rates were those used in Experiment 2. The display was generated using the same procedures described in Experiment 1 except that the intensity parameter for the flickering regions was varied to give the following luminance differences between the flickering and non-flickering regions (defined by the ratio flickering/non-flickering): .75, 1, 1.25, 1.5, 1.75, 2. When the ratio was 1, the average temporal luminance between the flickering and non-flickering areas was matched; when the ratio was 1.5, the average temporal luminance of the flickering region was one and a half times that of the average temporal luminance of the non-flickering regions. Square-wave flicker was used throughout and the amplitude of modulation was always set at 100%. As in the other experiments, the display was exposed for 10 secs. in each trial.

Procedure. Viewing conditions were similar to those of Experiments 1 and 2. Combinations of bar-widths and temporal frequency, and luminance ratios between flickering and non-flickering regions were presented in randomized order. Data were collected using the method described in the other experiments except that in addition to making region-segregation and depth judgments, the subject also judged whether the flickering areas appeared dimmer, brighter, or were equally as bright as the non-flickering regions.

Results

First, no subject perceived the flickering regions in front of the non-flickering regions for all the luminance ratios across the bar widths and temporal frequencies used in this study (binomial, p. < .001). All subjects also perceived the flickering and non-flickering dots segregated into distinct perceptual regions regardless of whether depth separation was seen between the two segregated areas (binomial, p. < .001).

The effect of a luminance difference between the flickering and non-flickering fields on the amount of perceived depth separation between the regions is shown in Figure 4. Data for each luminance ratio (flickering/

Insert Figure 4 about here

---

non-flickering) are shown in the six panels.

The tuning functions shown in Figure 4 were generated with the procedure described in Experiment 1. Five clusters emerged from the depth judgments made by the subjects. These are represented in the figure as iso-depth contours connecting temporal frequencies and bar widths giving similar amount of depth segregation between the flickering and non-flickering regions. Within each contour, no subject's depth judgment fell outside that contour.

When the flickering regions had an average temporal luminance .75 times that of the non-flickering regions' temporal luminance, the depth separation between the two areas was greatest (see Panel A). Substantial amount of depth was also seen across all the temporal and spatial frequencies used. In this condition, six out of seven observers perceived the flickering regions to be dimmer than the non-flickering regions. Only one observer judged all the regions to be equally bright. Not much difference existed between the spatio-temporal tuning functions at luminance ratios 1, 1.25, 1.5. (Panels B, C, D in Figure 4).

When the ratio was 1, all the observers perceived all the regions as equally bright at the lower temporal frequencies. Two subjects judged the flickering regions to be brighter than the non-flickering regions at the high temporal frequencies (above 7.1 Hz.). At luminance ratios 1.25 and 1.5, all observers reported the flickering regions as brighter than the non-flickering regions although the latter were perceived to stand out in front. When the luminance of the flickering regions were 1.75 and 2 times that of the non-flickering regions, the amount of perceived depth separation between these areas diminished. At bar widths and temporal frequencies where depth was maximum in the original (equal luminance) condition, the magnitude of depth now dropped by an average of 1.5 cm. When the luminance ratio was 1.75, sizable depth was only perceived at a bar-width of 1.0 deg. across 5.5, 6.3, 7.1, and 8.3 Hz. At the other spatio-temporal regions, perceived depth segregation was weak. At a luminance ratio of 2, the amount of perceived depth was minimal across all the bar-widths and temporal frequencies. Maximum depth perceived here amounted to no more than 1.7 cm., compared to at least 3.8 cm. in that same bar width and temporal frequency range for the smaller luminance ratios (1, 1.25, 1.5).

A three-way ANOVA (repeated measures) showed all main effects to be significant (bar width factor: $F = 16.81$, $df = (3,18)$, $p. < .001$; temporal frequency factor: $F = 14.64$, $df = (4,24)$, $p. < .001$; luminance ratio: $F = 4.86$, $df = (5,30)$, $p. < .005$). All the two-way interactions were significant (bar width x temporal frequency: $F = 2.94$, $df = (12,72)$, $p. < .005$; bar width x luminance ratio: $F = 2.88$, $df = (18,90)$, $p. < .005$; temporal frequency x luminance ratio: $F = 1.52$, $df = (30,120)$, $p. < .01$). The three-way interaction tended to approach significance but did not reach the statistical criterion ($F = 0.86$, $df = (12, 270)$).

A post-hoc comparison between means for a factorial design was performed on the luminance ratio factor. We decided to compare means within category rather than to make comparisons between specific cells although the interactions in the ANOVA were significant. This is because the issue of interest here is to assess the contribution of particular luminance ratios to the overall effect of luminance difference on the perceived amount of depth separation between the flickering and non-flickering regions. A comparison of specific cells would fail to highlight this point. Means for the luminance ratios .75, 1.75, and 2 were all significantly different from the rest of the ratios. No significant differences were found among the means for the ratios 1, 1.25, 1.5.

We also wanted to know whether perceived brightness of the flickering areas is related to the magnitude of perceived depth separation between the flickering and non-flickering regions especially when the luminance ratio was held constant. We therefore correlated the depth judgment (indexed by adjustment of a line's length) and the brightness comparison made between the flickering and non-flickering areas (indexed by three categorical responses of brighter, dimmer, and same). A multiple-covariate ANCOVA was performed on the data. Magnitude of perceived depth was correlated with the independent variable brightness judgment, when bar width, temporal frequency, and luminance ratio are covariates. The F ratio was found to be significant at the .05 level ($F = 3.78$, $df = 2,15$).

It should be noted that at the luminance ratios of 1.75 and 2, although the magnitude of depth separation between the non-flickering and flickering areas diminished and the former was perceived as bright, no observer saw the flickering regions in front of the non-flickering regions.

General Discussion

Temporal Tuning and Bar Width

Flickering a region of a filled visual field segregates it in depth behind areas which are not flickered. This depth segregation was found to be strongly dependent on temporal frequency and width of the bars in our display, thus establishing it as an effect related to temporal frequency response rather than merely to grouping by flicker. This interpretation is also supported by the observation that segregation of the flickering and non-flickering dots into distinct perceptual regions was perceived at bar width and temporal frequencies where no depth was seen.

The tuning functions obtained in Experiment 1 show that depth segregation produced by flicker is optimal between bar widths of 1.35 deg. and .68 deg., and between temporal frequencies of 6.3 Hz. and 8.3 Hz. These tuning characteristics resemble those of visual channels responding maximally to high temporal frequency and low spatial frequency (Robson, 1966; Tolhurst, 1975; Breitmeyer & Ganz, 1977; Legge, 1978; Burbeck & Kelly, 1981).

Modulation Amplitude and Perceived Depth

The results of Experiment 2 show that the magnitude of the depth effect is affected by the amplitude of the temporal modulation. It is at its maximum at 100% and diminishes as percent modulation decreases. This is no surprise since one would expect the response strength to diminish as contrast in the temporal dimension decreases. However, the analyses of interactions between modulation amplitude, bar width, and temporal frequencies suggest that not only the overall response strength was dampened by reduction of modulation amplitude, but the gradient of the iso-depth contours were also altered (reflected in the interactions between bar width and temporal frequency, and modulation amplitude). As depth of temporal modulation decreased, the iso-depth gradients became gentler at the lower temporal and higher temporal frequencies. Iso-depth contour gradients at the high temporal frequency and small bar width range were relatively unaffected although the strength of the depth response was dampened. (This result is reminiscent of the findings of Keck, Palella, & Pantle (1976), Burbeck & Kelly (1981) who showed that high temporal frequency channels saturate at low contrast.)

Apparent Brightness and Magnitude of Perceived Depth

Luminance differences between the flickering and non-flickering regions also determines the amount of depth separation seen between these areas. However, this occurred only when the average luminance of the flickering regions was lower or substantially higher than that of the non-flickering regions. This is also not surprising since regions with a high luminance level were also perceived as subjectively brighter, and it is known that subjective brightness of a stimulus affects its perceived depth from the observer (Ittelson, 1960). Dimmer objects are judged as farther away than bright objects. Perceived brightness as a cue to depth may counteract the depth effect produced by flicker when the flickering areas are subjectively brighter than the non-flickering areas, and enhancing depth when the reverse is true. It is important to note, though, that subjective brightness cannot be the explanation for the depth segregation perceived since flickering regions which were judged brighter were still localized behind the dimmer non-flickering areas.

Depth, Figure-ground, and Spatio-temporal Response

There is much evidence that channels tuned to high temporal and low spatial frequencies also respond best to a changing stimulus, while channels tuned to low temporal and high spatial frequencies respond maximally to a steadily-presented stimulus (Kulikowski & Tolhurst, 1973; Tolhurst, 1975; Breitmeyer & Ganz, 1977; Legge, 1978). There have been attempts to link the two sets of visual channels (those most sensitive to high temporal and low spatial frequencies and those most sensitive to low temporal and high spatial frequencies) to flicker/motion and pattern perception respectively (Breitmeyer & Ganz, 1976; von Grunau, 1978). This flicker-pattern dichotomy is supported by findings showing threshold differences between pattern and flicker and by the greater sensitivity of the visual system to flicker and motion at the low end of the spatial frequency spectrum (Keesey, 1972; Tolhurst, 1973; King-Smith & Kulikowski, 1975; Burbeck, 1971).

Our findings suggest a relationship between depth segregation and visual channels sensitive to low temporal/high spatial frequencies and high temporal/ low spatial frequencies. Our findings also suggest a relationship between flicker-induced depth segregation and figure-ground perception. It has been demonstrated that figural regions in the visual field are often perceived in front of their background (Rubin, 1922; Koffka, 1953; Hochberg, 1971; Coren, 1973; Kaniza, 1979). Moreover, perceptual regions which stand in front of the rest of the visual field are often perceived as figures (Julesz, 1971). In our experiments, observers often described the non-flickering regions as "bars emerging out in front of a flickering background". Our finding that non-flickering areas of the visual field are perceptually localized in front of the flickering areas can be defined as another instance of figure-ground segregation.

The resemblance between figure-ground segregation and the depth separation we found may go further than definition. Recently it has been proposed that figure and ground perception involve different visual processes (Julesz, 1978; Calis & Leeuwenberg, 1981; Wong & Weisstein, 1982). There is growing evidence that while figure analysis uses information from the high spatial frequencies and is specialized in detail resolution of contour, displacement, and orientation (Weitzman, 1963; Bridgeman, 1981; Wong & Weisstein, 1982) ground analysis is most sensitive to the global properties of a pattern contained in the low spatial frequencies. This was shown by an earlier finding of ours: viz. that high spatial frequency targets are enhanced in figural regions while low spatial frequency targets are enhanced in ground regions (Wong & Weisstein, 1982).

Given the results of the experiments reported here and given the possible links between figure and ground analyzing processes and the different spatio-temporal channels, it is tempting to look at our experimental manipulation as activating the figure and ground analyzing processes. Channels tuned to high temporal/low spatial frequencies responding to flicker would "signal" ground and channels responding maximally to low temporal/high spatial frequencies would "signal" figure. The mounting evidence of a more than casual association between a high temporal/low spatial frequency response and figure analysis and low spatial and high temporal frequency response and ground analysis makes it reasonable to think that the processes underlying flicker-induced depth and those mediating figure-ground segregation are closely related.

References

Bridgeman, B. Cognitive factors in subjective stabilization of the visual world. Acta Psychologia, 1981, 48, 111-121.

Breitmeyer, B., & Ganz, L. Implications of sustained and transient channels for theories of visual pattern masking, saccadic suppression, and information processing. Psychological Review, 1976, 83, 1-36.

Breitmeyer, B., & Ganz, L. Temporal studies with flashed gratings: Inferences about human transient and sustained channels. Vision Research, 1977, 17, 861-865.

Burbeck, C. A. Criterion-free pattern and flicker thresholds. Journal of the Optical Society of America, 1981, 71, 1343-1350.

Burbeck, C. A., & Kelly, D. H. Contrast gain measurement and the transient/sustained dichotomy. Journal of the Optical Society of America, 1981, 71, 1335-1342.

Johansson, G. Configurations in event perception. Uppsala: Almkvist and Wiskell, 1950.

Calis, G., & Leeuwenberg, E. L. J. Grounding the figure. Journal of Experimental Psychology: Human Perception and Performance, 1981, 7, 1386-1397.

Coren, S. Subjective contours and apparent depth. Psychological Review, 1972, 79, 359-367.

Hochberg, J. Color and shape. In L. A. Riggs and J. F. Kling (Eds.), Woodsworth and Scholsberg's Experimental Psychology, 3rd edition. New York: Holt, Reinhart and Winston, 1972.

Ittelson, W. H. Visual space perception. New York: Springer, 1960.

Julesz, B. The foundations of cyclopean perception. Chicago: University of Chicago Press, 1971.

Julesz, B. Perceptual limits of texture discrimination and their implications to figure-ground separation. In E. L. J. Leeuwenberg and H. F. J. M. Buffart (Eds), Formal theories of perception. New York: Wiley, 1978.

Kaniza, G. Organization in vision. New York: Praeger, 1979.

Keck, M. J., Palella, T. D., & Pantle, A. Motion aftereffect as a function of the contrast of sinusoidal gratings. Vision Research, 1976, 16, 187-192.

Keesey, U. T. Flicker and pattern detection: A comparison of thresholds. Journal of the Optical Society of America, 1972, 62, 446-448.

King-Smith, P. E., & Kulikowski, J. J. Pattern and flicker detection analyzed by subthreshold summation. Journal of Physiology, 1975, 519-548.

Koffka, K. Principles of Gestalt psychology. New York: Harcourt-Brace, 1935.

Kulikowski, J. J. Psychophysical evidence for sustained and transient detectors in human vision. Journal of Physiology, 1973, 232, 777-784.

Legge, G. E. Sustained and transient mechanisms in human vision: temporal and spatial properties. Vision Research, 1978, 18, 69-81.

Robson, J. G. Spatial and temporal contrast sensitivity functions of the visual system. Journal of the Optical Society of America, 1966, 56, 1411-1412.

Scher, A., Schneier, M., & Rosenfeld, A. Clusteringof collinear line segments. Pattern Recognition, 1982, 15, 85-92.

Tolhurst, D. J. Separate channels for the analysis of the shape and the movement of a moving visual stimulus. Journal of Physiology, 1973, 231, 385-402.

von Granau, M. W. Dissociation and inhibition of form and motion information in the human visual system. Vision Research, 1978, 18, 1485-1490.

Weitzman, B. A threshold difference produced by a figure-ground dichotomy. Journal of Experimental Psychology, 1963, 66, 201-205.

Wertheimer, M. Untersuchunger zur Lehre von der Gestalt, II. Psychologische Forschung, 1923, 4, 301-350.

Wong, E., & Weisstein, N. A new context-superiority effect: Line segments are more visible against a figure than against a ground. Science, in press.

Wong, E., & Weisstein, N. Sharp targets are detected better against a figure and blurred targets are detected better against a background. Paper submitted for publication.

Wong. E., & Weisstein, N. A new phenomenon in perceptual flickering fields are localized in a depth plane behind non-flickering fields. Paper submitted for publication.

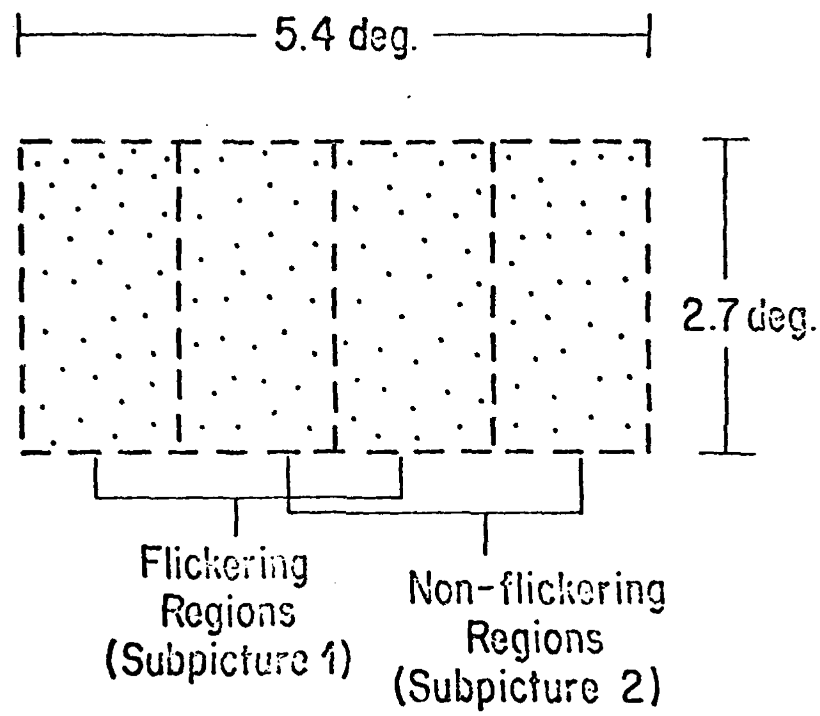
Figure Captions

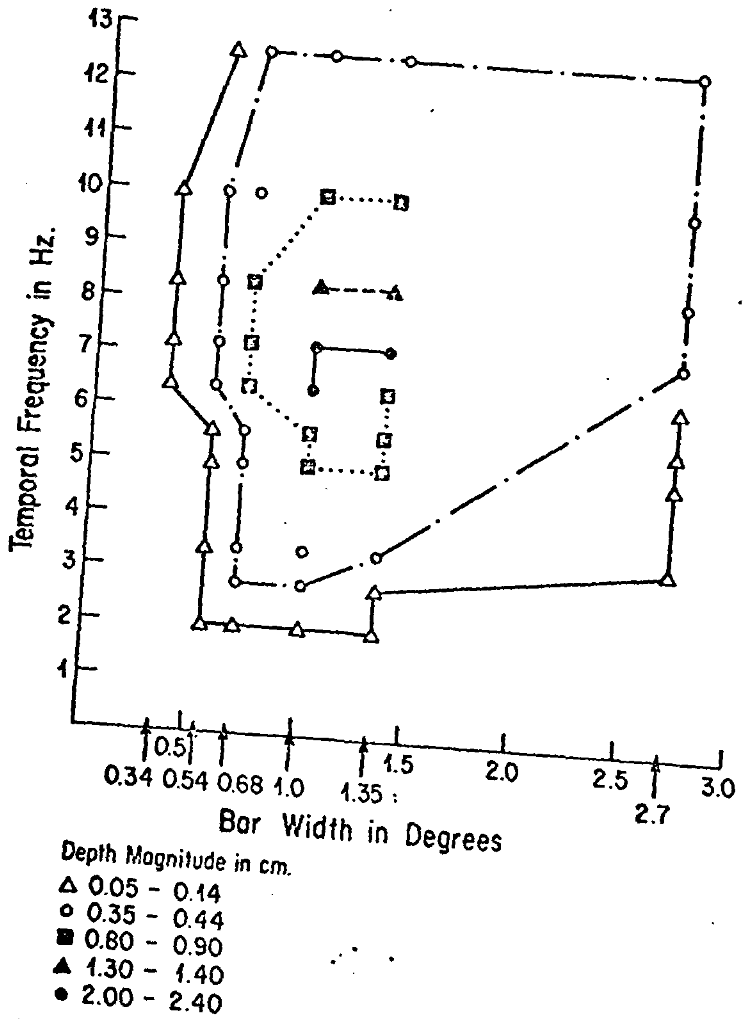
Figure 1. Illustration of the display used in the experiments. This is an example of flickering and non-flickering bars composed of random-dots with a bar width of .68 deg.

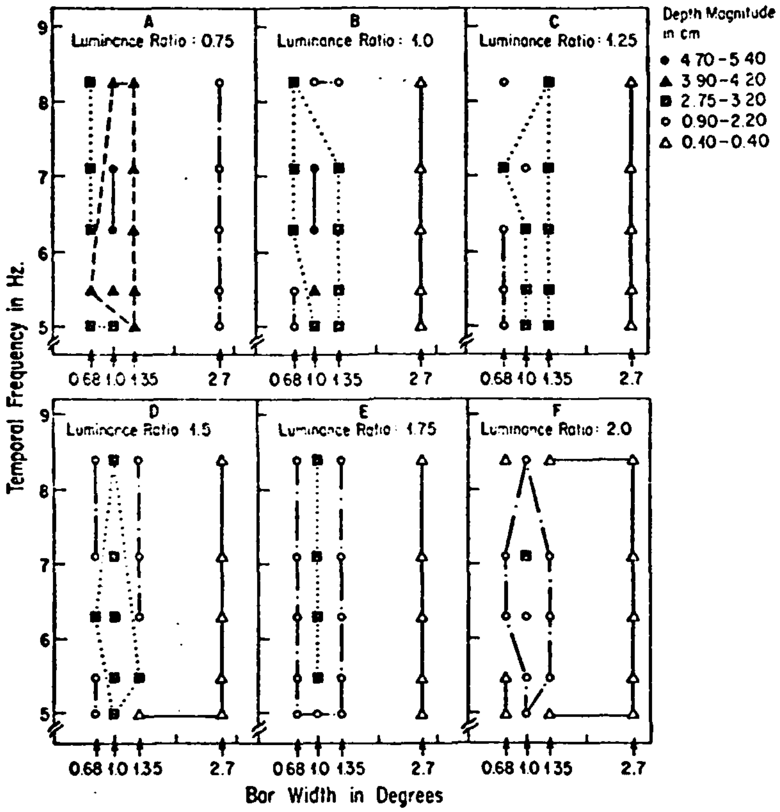
Figure 2. Tuning function for depth segregation induced by flicker. Contours connect bar width and temporal frequency loci where similar amount of depth is perceived.

Figure 3. Comparison of tuning functions of the depth effect induced by flicker at temporal modulation amplitudes of 100%, 75%, 50%, and 25%.

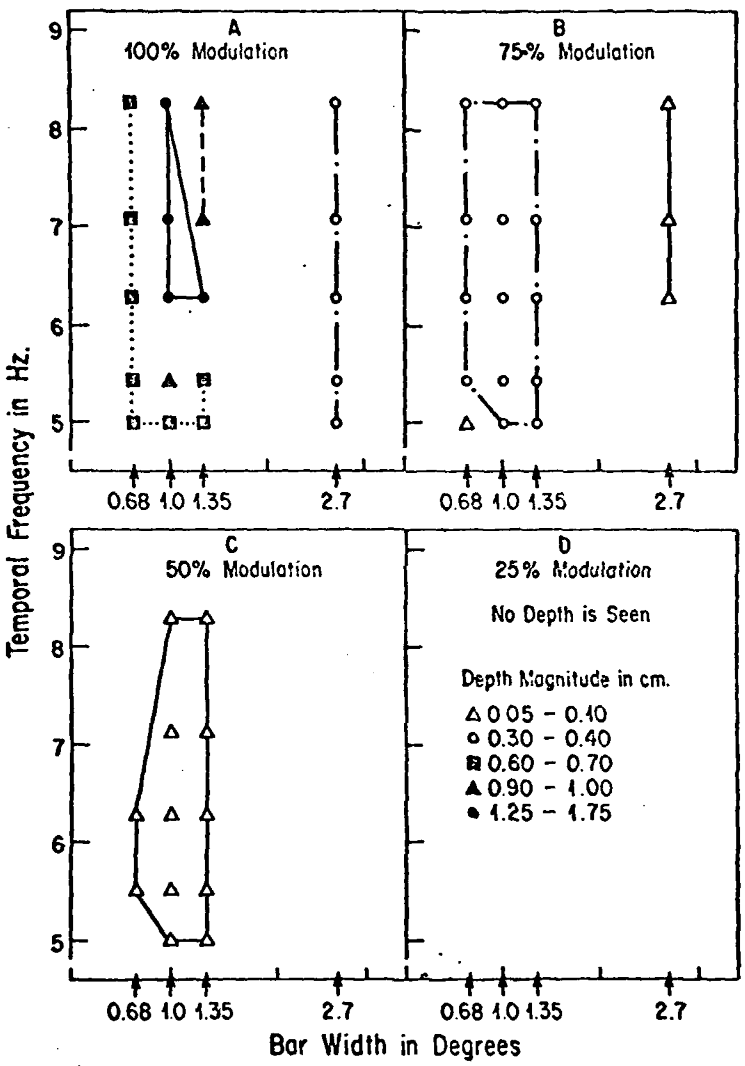
Figure 4. Comparison of tuning functions of the depth effect induced by flicker at six luminance differences between the flickering and non-flickering regions. The ratios are expressed as flickering areas/non-flickering areas.


5.4 deg.
2.7 deg.
Flickering
Regions
(Subpicture 1)
Non-flickering
Regions
(Subpicture 2)


Temporal Frequency in Hz.
13
12
11
10
9
8
7
6
5
4
3
2
1
0.5
1.5
2.0
2.5
3.0
0.34
0.54
0.68
1.0
1.35
2.7
Bar Width in Degrees
Depth Magnitude in cm.
△ 0.05 - 0.14
○ 0.35 - 0.44
■ 0.80 - 0.90
▲ 1.30 - 1.40
● 2.00 - 2.40



A
Luminance Ratio : 0.75
B
Luminance Ratio : 1.0
C
Luminance Ratio : 1.25
D
Luminance Ratio 1.5
E
Luminance Ratio 1.75
F
Luminance Ratio : 2.0
Depth Magnitude in cm
4.70-5.40
3.90-4.20
2.75-3.20
0.90-2.20
0.40-0.40
Temporal Frequency in Hz.
Bar Width in Degrees
0.68 1.0 1.35 2.7



A
100% Modulation
B
75% Modulation
C
50% Modulation
D
25% Modulation
No Depth is Seen
Depth Magnitude in cm.
△ 0.05 - 0.10
○ 0.30 - 0.40
▣ 0.60 - 0.70
▲ 0.90 - 1.00
● 1.25 - 1.75
Temporal Frequency in Hz.
Bar Width in Degrees
0.68 1.0 1.35 2.7